

# On the Satisfiability of Quantum Circuits of Small Treewidth <sup>\*</sup>

Mateus de Oliveira Oliveira

**Abstract** It has been known for almost three decades that many NP-hard optimization problems can be solved in polynomial time when restricted to structures of constant treewidth. In this work we provide the first extension of such results to the quantum setting. We show that given a quantum circuit  $C$  with  $n$  uninitialized inputs,  $\text{poly}(n)$  gates, and treewidth  $t$ , one can compute in time  $(\frac{n}{\delta})^{\exp(O(t))}$  a classical assignment  $y \in \{0, 1\}^n$  that maximizes the acceptance probability of  $C$  up to a  $\delta$  additive factor. In particular, our algorithm runs in polynomial time if  $t$  is constant and  $1/\text{poly}(n) < \delta < 1$ . For unrestricted values of  $t$ , this problem is known to be complete for the complexity class QCMA, a quantum generalization of MA. In contrast, we show that the same problem is NP-complete if  $t = O(\log n)$  even when  $\delta$  is constant.

On the other hand, we show that given a  $n$ -input quantum circuit  $C$  of treewidth  $t = O(\log n)$ , and a constant  $\delta < 1/2$ , it is QMA-complete to determine whether there exists a quantum state  $|\varphi\rangle \in (\mathbb{C}^d)^{\otimes n}$  such that the acceptance probability of  $C|\varphi\rangle$  is greater than  $1 - \delta$ , or whether for every such state  $|\varphi\rangle$ , the acceptance probability of  $C|\varphi\rangle$  is less than  $\delta$ . As a consequence, under the widely believed assumption that  $\text{QMA} \neq \text{NP}$ , we have that quantum witnesses are strictly more powerful than classical witnesses with respect to Merlin-Arthur protocols in which the verifier is a quantum circuit of logarithmic treewidth.

**Keywords** Treewidth · Satisfiability of Quantum Circuits · Tensor Networks · Merlin-Arthur Protocols

---

<sup>\*</sup> This is an extended version of a paper that appeared at CSR 2015 [14].

This work was supported by the European Research Council, ERC grant agreement 339691, within the context of the project Feasibility, Logic and Randomness (FEALORA).

---

M. de Oliveira Oliveira  
Institute of Mathematics - Czech Academy of Sciences  
Žitná 25, CZ - 115 67 Praha 1, Czech Republic  
E-mail: mateus.oliveira@math.cas.cz

## 1 Introduction

The notions of tree decomposition and treewidth of a graph [22] play a central role in algorithmic theory. On the one hand, many natural classes of graphs have small treewidth. For instance, trees have treewidth at most 1, series-parallel graphs and outer-planar graphs have treewidth at most 2, Halin graphs have treewidth at most 3, and  $k$ -outerplanar graphs for fixed  $k$  have treewidth  $O(k)$ . On the other hand, many problems that are hard for NP on general graphs, and even problems that are hard for higher levels of the polynomial hierarchy, may be solved in polynomial time when restricted to graphs of constant tree-width [5, 6, 13]. In particular, during the last decade, several algorithms running in time  $2^{O(t)} \cdot n^{O(1)}$  have been proposed for the satisfiability of classical circuits<sup>1</sup> and boolean constraint satisfaction problems of size  $n$  and treewidth  $t$  [3, 4, 12, 15].

In this work, we identify for the first time a natural quantum optimization problem that becomes feasible when restricted to graphs of constant treewidth. More precisely, we show how to find in polynomial time a classical assignment that maximizes, up to an inverse polynomial additive factor, the acceptance probability of a quantum circuit of constant treewidth. For quantum circuits of unrestricted treewidth this problem is complete for QCMA, a quantum generalization of MA [2]. Before stating our main result, we fix some notation. If  $C$  is a quantum circuit acting on  $n$   $d$ -dimensional qudits, and  $|\psi\rangle$  is a quantum state in  $(\mathbb{C}^d)^{\otimes n}$ , then we denote by  $Pr(C, |\psi\rangle)$  the probability that the state of the output of  $C$  collapses to  $|1\rangle$  when the input of  $C$  is initialized with  $|\psi\rangle$  and the output is measured in the standard basis  $\{|0\rangle, |1\rangle, \dots, |d-1\rangle\}$ . If  $y$  is a string in  $\{0, \dots, d-1\}^n$  then we let  $|y\rangle = \otimes_{i=1}^n |y_i\rangle$  denote the basis state corresponding to  $y$ . We let  $Pr^{cl}(C) = \max_{y \in \{0, \dots, d-1\}^n} Pr(C, |y\rangle)$  denote the maximum acceptance probability of  $C$  among all classical input strings in  $\{0, \dots, d-1\}^n$ . The treewidth of a quantum circuit is defined as the treewidth of its underlying undirected graph.

**Theorem 1 (Main Theorem)** *Let  $C$  be a quantum circuit with  $n$  uninitialized inputs,  $\text{poly}(n)$  gates, and treewidth  $t$ . For each  $\delta$  with  $1/\text{poly}(n) < \delta < 1$ , one may find in time  $(\frac{n}{\delta})^{\exp(O(t))}$  a string  $y \in \{0, \dots, d-1\}^n$  such that  $|Pr(C, |y\rangle) - Pr^{cl}(C)| \leq \delta$ .*

We note that the algorithm that finds the string  $y \in \{0, 1\}^n$  in Theorem 1 is completely deterministic. The use of treewidth in quantum algorithmics was pioneered by Markov and Shi [20] who showed that quantum circuits of logarithmic treewidth can be simulated in polynomial time with exponentially high precision. Note that the simulation of quantum circuits [16, 18, 20, 25, 26] deals with the problem of computing the acceptance probability of a quantum circuit when all inputs are already initialized, and thus may be regarded as a generalization of the classical P-complete problem CIRCUIT-VALUE. On the other hand, Theorem 1 deals with the problem of finding a classical assignment that maximizes the acceptance probability of a quantum circuit with uninitialized inputs, and thus may be regarded as a generalization of the classical NP-complete problem CIRCUIT-SAT. In this sense, Theorem 1 is the first result showing that a quantum generalization of CIRCUIT-SAT can be solved in polynomial time when restricted to circuits of constant treewidth.

---

<sup>1</sup> In the case of classical circuits, it is assumed that each variable labels a unique input of unbounded fan-out.

It is interesting to determine whether the time complexity of our algorithm can be substantially improved. To address this question, we first introduce the *online-width* of a circuit, a width measure for DAGs that is at least as large as the treewidth of their underlying undirected graphs. If  $G = (V, E)$  is a directed graph and  $V_1, V_2 \subseteq V$  are two subsets of vertices of  $V$  with  $V_1 \cap V_2 = \emptyset$  then we let  $E(V_1, V_2)$  be the set of all edges with one endpoint in  $V_1$  and another endpoint in  $V_2$ . If  $\omega = (v_1, v_2, \dots, v_n)$  is a total ordering of the vertices in  $V$ , then we let  $\mathbf{cw}(G, \omega) = \max_i |E(\{v_1, \dots, v_i\}, \{v_{i+1}, \dots, v_n\})|$ . The *cutwidth* of  $G$  is defined as  $\mathbf{cw}(G) = \min_{\omega} \mathbf{cw}(G, \omega)$  where the minimum is taken over all possible total orderings of the vertices of  $G$  [24]. If  $G$  is a DAG, then the *online-width* of  $G$  is defined as  $\mathbf{ow}(G) = \min_{\omega} \mathbf{cw}(G, \omega)$  where the minimum is taken only among the *topological orderings* of  $G$ . Treewidth, cutwidth and online-width are compared as follows<sup>2</sup>.

$$\mathbf{tw}(G) \leq \mathbf{cw}(G) \leq \mathbf{ow}(G) \quad (1)$$

Theorem 2 below states that finding a classical assignment that maximizes the acceptance probability of a quantum circuit of logarithmic online-width is NP-complete even when  $\delta$  is constant. We note that the same completeness result holds with respect to circuits of logarithmic treewidth.

**Theorem 2** *For any constant  $\delta$  with  $0 < \delta < 1$ , the following problem is NP-complete: Given a quantum circuit  $C$  of online-width  $O(\log n)$  with  $n$  uninitialized inputs and  $\text{poly}(n)$  gates, determine whether  $\text{Pr}^{cl}(C) = 1$  or whether  $\text{Pr}^{cl}(C) \leq \delta$ .*

An analog completeness result holds when the verifier is restricted to have logarithmic online-width and the witness is allowed to be an arbitrary quantum state. It was shown by Kitaev [19] that finding a  $\delta$ -optimal quantum witness for a quantum circuit of unrestricted width is complete for the complexity class QMA for any constant  $\delta$ . Interestingly, Kitaev's completeness result is preserved when the quantum circuits are restricted to have logarithmic online-width. If  $C$  is a quantum circuit with  $n$  inputs, then we let  $\text{Pr}^{qu}(C) = \max_{|\psi\rangle} \text{Pr}(C, |\psi\rangle)$  be the maximum acceptance probability among all  $n$ -qudit quantum states  $|\psi\rangle$ .

**Theorem 3** *For any constant  $\delta$  with  $0 < \delta < 1/2$ , the following problem is QMA-Complete: Given a quantum circuit  $C$  of online-width  $O(\log n)$  with  $n$  uninitialized inputs and  $\text{poly}(n)$  gates, determine whether  $\text{Pr}^{qu}(C) \geq 1 - \delta$  or whether  $\text{Pr}^{qu}(C) \leq \delta$ .*

We analyse the implications of theorems 2 and 3 to quantum generalizations of Merlin-Arthur protocols. In this setting, Arthur, a polynomial sized quantum circuit, must decide the membership of a string  $x$  to a given language  $\mathcal{L}$  by analysing a quantum state  $|\psi\rangle$  provided by Merlin. In the case that  $x \in \mathcal{L}$ , there is always a quantum state  $|\psi\rangle$  that is accepted by Arthur with probability at least  $2/3$ . On the other hand, if  $x \notin \mathcal{L}$  then no state is accepted by Arthur with probability greater than  $1/3$ . The class of all languages that can be decided via some quantum Merlin-Arthur protocol is denoted by QMA. The importance of QMA stems from the fact that this class has several natural complete problems [11, 19]. Additionally, the oracle version of QMA contains problems, such as the group non-membership problem [27] which are provably not in the oracle version of MA and hence not in the oracle version of NP [7]. The class QCMA is defined analogously, except for the fact that the witness provided by Merlin is a product state encoding a classical string. Below we define width parameterized versions of QMA.

<sup>2</sup> A proof that  $\mathbf{tw}(G) \leq \mathbf{cw}(G)$  can be found in [8].

**Definition 1** A language  $\mathcal{L} \subseteq \{0,1\}^*$  belongs to the class  $\text{QMA}[\text{tw}, f(n)]$  if there exists a polynomial time constructible family of quantum circuits  $\{C_x\}_{x \in \{0,1\}^*}$  such that for every  $x \in \{0,1\}^*$ ,  $C_x$  has treewidth at most  $f(|x|)$  and

- if  $x \in \mathcal{L}$  then there exists a quantum state  $|\psi\rangle$  such that  $C_x$  accepts  $|\psi\rangle$  with probability at least  $2/3$ ,
- if  $x \notin \mathcal{L}$  then for each quantum state  $|\psi\rangle$ ,  $C_x$  accepts  $|\psi\rangle$  with probability at most  $1/3$ .

The class  $\text{QCMA}[\text{tw}, f(n)]$  is defined analogously, except that the witness  $|y\rangle$  is required to be the basis state encoding of a classical string  $y$ .

Definition 1 can be extended naturally to other width measures such as online-width. For instance,  $\text{QMA}[\text{ow}, f(n)]$  and  $\text{QCMA}[\text{ow}, f(n)]$  denote the classes of languages that can be decided by quantum Merlin-Arthur games with respectively quantum and classical witnesses, in which the verifier is required to have online-width at most  $f(n)$ . We note that the classes  $\text{QMA}$  and  $\text{QCMA}$  can be defined respectively as  $\text{QMA}[\text{ow}, \text{poly}(n)]$  and  $\text{QCMA}[\text{ow}, \text{poly}(n)]$ , since the online-width of a circuit can be at most quadratic in its number of gates. In the next corollary we analyse the complexity of low-width quantum Merlin-Arthur protocols with classical and quantum witnesses.

**Corollary 1**

- i.  $\text{QCMA}[\text{tw}, O(1)] \subseteq P$ .
- ii.  $\text{QCMA}[\text{tw}, O(\log n)] = \text{QCMA}[\text{ow}, O(\log n)] = \text{NP}$ .
- iii.  $\text{QMA}[\text{tw}, O(\log n)] = \text{QMA}[\text{ow}, O(\log n)] = \text{QMA}$ .

We note that Corollary 1.i is a consequence of Theorem 1, Corollary 1.ii is a consequence of Theorem 2, and Corollary 1.iii is a consequence of Theorem 3. Under the plausible assumption that  $\text{QMA} \neq \text{NP}$ , Corollary 1 implies that whenever Arthur is restricted to be a quantum circuit of logarithmic treewidth, quantum Merlin-Arthur protocols differ in power with respect to whether the witness provided by Merlin is classical or quantum. We observe that obtaining a similar separation between the power of classical and quantum witnesses when Arthur is allowed to be a quantum circuit of polynomial treewidth is equivalent to determining whether  $\text{QMA} \neq \text{QCMA}$ . This question remains widely open.

## 1.1 Organization of the Paper

In Section 2 we will define basic notions such as quantum circuits, tree decompositions and treewidth. Sections 3 to 7 will be dedicated to the proof of our main theorem (Theorem 1). The proof of this theorem will be sketched in Subsection 1.2. In Section 8 we will prove Theorem 2, and in Section 9 we will prove Theorem 3. We will conclude this paper by making some final considerations and by stating some open problems in Section 10.

## 1.2 Sketch of the Proof of Theorem 1

We will prove Theorem 1 using a combination of techniques from tensor network theory, structural graph theory and dynamic programming. We will start by introducing in Section 3 the notion of *abstract network*. Intuitively, an abstract network

is a list  $\mathcal{N} = \{\mathcal{I}_1, \dots, \mathcal{I}_n\}$  of finite subsets of positive integers, called *index sets*. Such an abstract network  $\mathcal{N}$  can be naturally associated with a graph  $G(\mathcal{N})$ . This graph is obtained by creating a vertex  $v_{\mathcal{I}}$  for each index set  $\mathcal{I} \in \mathcal{N}$  and by adding  $k$  edges between two vertices  $v_{\mathcal{I}}$  and  $v_{\mathcal{I}'}$  if and only if  $|\mathcal{I} \cap \mathcal{I}'| = k$ . We will be interested in the process of contracting the vertices of the graph  $G(\mathcal{N})$  into a single vertex. Such a contraction process will be represented by a data structure called *contraction tree*. The complexity of a contraction tree will be measured via two parameters: its *rank*, and its *height*. In Section 4 we will show that if the graph  $G(\mathcal{N})$  associated with an abstract network  $\mathcal{N}$  has treewidth  $t$  and maximum degree  $\Delta$ , then one can efficiently construct a contraction tree for  $\mathcal{N}$  of rank  $O(\Delta \cdot t)$  and height  $O(\Delta \cdot t \cdot \log |\mathcal{N}|)$  (Theorem 4). As we will argue below, Theorem 4 will play an important role in the proof of our main result.

Abstract networks can be used to define both the well known notion of *tensor network* (Section 5), and the new notion of *feasibility tensor network* (Section 6). Within this formalism, a tensor network can be viewed as a pair  $(\mathcal{N}, \lambda)$  where  $\lambda$  is a function that associates a tensor  $\lambda(\mathcal{I})$  of rank  $|\mathcal{I}|$  with each index set  $\mathcal{I}$  of  $\mathcal{N}$ . One can define a notion of contraction for tensor networks with basis on the notion of contraction for abstract networks. Contracting a tensor network  $(\mathcal{N}, \lambda)$  yields a complex number  $g$ , i.e., a tensor of rank 0. The value of the tensor network, denoted by  $\text{val}(\mathcal{N}, \lambda)$ , is defined as the absolute value of  $g$ . It can be shown that the problem of computing the acceptance probability of a quantum circuit  $C$  in which all inputs are initialized can be reduced to the problem of computing the value of a suitable tensor network  $(\mathcal{N}_C, \lambda_C)$ .

On the other hand, a feasibility tensor network is a pair  $(\mathcal{N}, A)$  where  $\mathcal{N}$  is an abstract network and  $A$  is a function that associates with each index set  $\mathcal{I} \in \mathcal{N}$  a *set* of tensors  $A(\mathcal{I})$  of rank  $|\mathcal{I}|$ . An initialization of  $(\mathcal{N}, A)$  is a function  $\lambda$  that associates a tensor  $\lambda(\mathcal{I}) \in A(\mathcal{I})$  with each index set of  $\mathcal{N}$ . Each such initialization yields a tensor network  $(\mathcal{N}, \lambda)$ . The value of the feasibility network  $(\mathcal{N}, A)$ , is defined as  $\text{VAL}(\mathcal{N}, A) = \max_{\lambda} \text{val}(\mathcal{N}, \lambda)$ , where  $\lambda$  ranges over all initializations of  $(\mathcal{N}, A)$ . As we will see in Section 6, the problem of finding a classical assignment that maximizes the acceptance probability of a quantum circuit with uninitialized inputs can be reduced to the problem of finding an initialization of maximum value for a suitable feasibility tensor network  $(\mathcal{N}_C, A_C)$ .

Let  $(\mathcal{N}, A)$  be a feasibility tensor network, and  $\varepsilon$  be a real number with  $0 < \varepsilon < 1$ . In Section 7 we will show that given a contraction tree for  $\mathcal{N}$  of rank  $r$  and height  $h$ , one can find in time  $|\mathcal{N}| \cdot \varepsilon^{-\exp(O(r \cdot \log d))}$  an initialization  $\lambda$  of  $(\mathcal{N}, A)$  such that  $|\text{val}(\mathcal{N}, \lambda) - \text{VAL}(\mathcal{N}, A)| \leq \varepsilon \cdot \exp(O(r \cdot h \cdot \log d))$  (Theorem 5). Therefore, to obtain a polynomial time algorithm for approximating the value of a feasibility tensor network  $(\mathcal{N}, A)$  up to a constant additive factor  $\delta$ , we need to keep the rank of the contraction tree bounded by a constant, and its height bounded by  $O(\log |\mathcal{N}|)$ .

As mentioned above, if the graph  $G(\mathcal{N})$  has treewidth  $t$  and maximum degree  $\Delta$ , then by Theorem 4 one can construct a contraction tree for  $\mathcal{N}$  of rank  $O(\Delta \cdot t)$ , and height  $O(\Delta \cdot t \cdot \log |\mathcal{N}|)$ . Therefore, by setting  $\varepsilon = \delta / |\mathcal{N}|^{O(\Delta^2 \cdot t^2 \cdot \log d)}$  we can use Theorem 5 to find in time  $(|\mathcal{N}|/\delta)^{\exp(O(\Delta \cdot t \cdot \log d))}$  an initialization  $\lambda$  for  $(\mathcal{N}, A)$  such that  $|\text{val}(\mathcal{N}, \lambda) - \text{val}(\mathcal{N}, A)| \leq \delta$  (Theorem 6).

Finally, let  $C$  be a quantum circuit with  $n$  uninitialized inputs, treewidth  $t$ , and  $n^{O(1)}$  gates drawn from a finite universal set of gates  $\mathcal{U}$ . Let  $(\mathcal{N}_C, A_C)$  be the feasibility tensor network associated with  $C$ . Then  $|\mathcal{N}| = n^{O(1)}$  and the graph  $G(\mathcal{N}_C)$  has treewidth  $t$ , and maximum degree bounded by a constant  $\Delta(\mathcal{U})$ . Therefore, as a

corollary of Theorem 6, a classical assignment that maximizes the acceptance probability of  $C$  up to a  $\delta$  additive factor can be found in time  $(n/\delta)^{\exp(O(\Delta(\mathcal{U}) \cdot t \cdot \log d))}$ . Since both  $\Delta(\mathcal{U})$  and the dimensionality  $d$  of the qudits over which  $C$  operates are constant, Theorem 1 follows.

## 2 Preliminaries

A  $d$ -dimensional qudit is a unit vector in the Hilbert space  $\mathcal{H}_d = \mathbb{C}^d$ . We fix an orthonormal basis for  $\mathcal{H}_d$  and label the vectors in this basis with  $|0\rangle, |1\rangle, \dots, |d-1\rangle$ . The  $n$ -fold tensor product of  $\mathcal{H}_d$  is denoted by  $\mathcal{H}_d^{\otimes n}$ . We denote by  $\mathbf{L}(\mathcal{H}_d^{\otimes n})$  the set of all linear operators on  $\mathcal{H}_d^{\otimes n}$ . An operator  $X$  on  $\mathbf{L}(\mathcal{H}_d^{\otimes n})$  is positive semidefinite if all its eigenvalues are non-negative. A density operator on  $n$  qudits is a positive semidefinite operator  $\rho \in \mathbf{L}(\mathcal{H}_d^{\otimes n})$  with trace  $\text{tr}(\rho) = 1$ . For a string  $y = y_1 y_2 \dots y_n \in \{0, 1, \dots, d-1\}^n$  we let  $\rho_y = \bigotimes_{i=1}^n |y_i\rangle\langle y_i|$  be the density operator of the state  $|y\rangle = \bigotimes_{i=1}^n |y_i\rangle$ . A map  $M : \mathbf{L}(\mathcal{H}_d^{\otimes q}) \rightarrow \mathbf{L}(\mathcal{H}_d^{\otimes r})$  is positive if  $M(\rho)$  is positive semidefinite whenever  $\rho$  is positive semidefinite. The map  $M$  is completely positive if the map  $I_k \otimes M$  is positive for every  $k \in \mathbb{N}$ , where  $I_k$  is the  $k \times k$  identity matrix. A quantum gate with  $q$  inputs and  $r$  outputs is a linear map  $Q : \mathbf{L}(\mathcal{H}_d^{\otimes q}) \rightarrow \mathbf{L}(\mathcal{H}_d^{\otimes r})$  that is completely positive, convex on density matrices, and such that  $0 \leq \text{tr}(Q(\rho)) \leq 1$  for any density matrix  $\rho$ . Linear maps satisfying these three properties formalize the notion of physically admissible quantum operation. We refer to [21] (Section 8.2.4) for a detailed discussion on physically admissible operations. A *positive-operator valued measure* (POVM) is a set  $\mathcal{X} = \{X_1, X_2, \dots, X_k\}$  of positive semidefinite operators such that  $\sum_i X_i = I$ . Each operator  $X_i$  in  $\mathcal{X}$  is called a *measurement element* of  $\mathcal{X}$ . If  $\mathcal{X}$  is a POVM then the probability of measuring outcome  $i$  after applying  $\mathcal{X}$  to  $\rho$  is given by  $\text{tr}(\rho X_i)$ . A single  $d$ -dimensional qudit measurement in the computational basis is defined as the POVM  $\mathcal{X} = \{|0\rangle\langle 0|, |1\rangle\langle 1|, \dots, |d-1\rangle\langle d-1|\}$ .

### 2.1 Quantum Circuits

We adopt the model of quantum circuits with mixed states introduced in [1]. Let  $\mathcal{U}$  be a finite universal set of quantum gates, and let  $\Delta(\mathcal{U})$  be the maximum number of inputs plus outputs of a gate in  $\mathcal{U}$ . A quantum circuit over  $\mathcal{U}$  is a connected directed acyclic graph<sup>3</sup>  $C = (V, E, \theta, \xi)$ , of maximum degree at most  $\Delta(\mathcal{U})$ , where  $V$  is a set of vertices,  $E$  a set of edges,  $\theta : V \rightarrow \mathcal{U}$  is a vertex labeling function and  $\xi : E \rightarrow \{1, \dots, |E|\}$  is an injective function that assigns a distinct number to each edge of  $C$ . The vertex set is partitioned into a set  $In$  (input vertices), a set  $Out$  (output vertices), and a set  $Mid = V \setminus (In \cup Out)$  (internal vertices). Each input vertex has in-degree 0 and out-degree 1, and each output vertex has in-degree 1 and out-degree 0. Each internal vertex has both in-degree and out-degree greater than 0. If  $v$  is an internal vertex with  $k$  incoming edges and  $l$  outgoing edges then  $v$  is labeled with a quantum gate  $\theta(v) \in \mathcal{U}$  with  $k$  inputs and  $l$  outputs. Each input vertex  $v$  is either labeled by  $\theta$  with an element from the set  $\{|0\rangle\langle 0|, |1\rangle\langle 1|, \dots, |d-1\rangle\langle d-1|\}$ , indicating that  $v$  is an initialized input, or with the symbol  $*$ , indicating that  $v$  is not

<sup>3</sup> All graphs in this work, being directed or undirected, may contain multiple edges, but no loops.

initialized. Finally, each output vertex  $v$  is labeled with an one-qudit measurement element  $\theta(v) \in \mathbf{L}(\mathcal{H}_d)$ . We let  $M(C) = \otimes_{v \in \text{Out}} \theta(v)$  denote the overall measurement element in  $\mathbf{L}(\mathcal{H}_d^{\otimes |\text{Out}(C)|})$  defined by  $C$ . A quantum circuit  $C$  with  $n$  uninitialized inputs and  $m$  outputs can be regarded as a superoperator  $C : \mathbf{L}(\mathcal{H}_d^{\otimes n}) \rightarrow \mathbf{L}(\mathcal{H}_d^{\otimes m})$ . If  $|\psi\rangle$  is a quantum state in  $\mathcal{H}_d^{\otimes n}$  then the acceptance probability of  $C$  when  $|\psi\rangle$  is assigned to the inputs of  $C$  is defined as  $\text{Pr}(C, |\psi\rangle) = \text{tr}[C(|\psi\rangle\langle\psi|) \cdot M(C)]$ .

## 2.2 Tree Decompositions and Treewidth

A tree is a connected acyclic graph  $T$  with set of nodes  $\text{nodes}(T)$  and set of arcs  $\text{arcs}(T)$ . A tree decomposition of a graph  $G = (V, E)$  consists of a pair  $(T, \beta)$  where  $T$  is a tree, and  $\beta : \text{nodes}(T) \rightarrow 2^V$  is a function that associates a set of vertices  $\beta(u)$  with each node  $u \in \text{nodes}(T)$ , in such a way that

- $\bigcup_{u \in \text{nodes}(T)} \beta(u) = V$ ,
- for every edge  $\{v, v'\} \in E$ , there is a node  $u \in \text{nodes}(T)$  such that  $\{v, v'\} \subseteq \beta(u)$ ,
- for every vertex  $v \in V$ , the set  $\{u \in \text{nodes}(T) \mid v \in \beta(u)\}$  induces a connected subtree of  $T$ .

The *width* of  $(T, \beta)$  is defined as  $\mathbf{w}(T, \beta) = \max_u \{|\beta(u)| - 1\}$ . The *treewidth*  $\mathbf{tw}(G)$  of a graph  $G$  is the minimum width of a tree decomposition of  $G$ .

If  $C = (V, E, \theta, \xi)$  is a quantum circuit, then the treewidth of  $C$  is defined as the treewidth of the undirected graph  $G_C = (V, E')$  obtained from  $C$  by forgetting vertex labels, edge labels, and direction of edges.

## 3 Abstract Networks

In this section we will introduce the notion of *abstract network*. In Section 5 we will use abstract networks to model the well known notion of tensor network, a formalism that is suitable for the simulation of quantum circuits. Subsequently, in Section 6, we will use abstract networks to define the new notion of *feasibility tensor network*, a formalism that is suitable for addressing the satisfiability of quantum circuits. Below, we call a possibly empty finite set  $\mathcal{I}$  of positive integers, an *index set*. We say that each number  $i$  in an index set  $\mathcal{I}$  is an *index*.

**Definition 2 (Abstract Network)** An *abstract network* is a finite list

$$\mathcal{N} = [\mathcal{I}_1, \dots, \mathcal{I}_m]$$

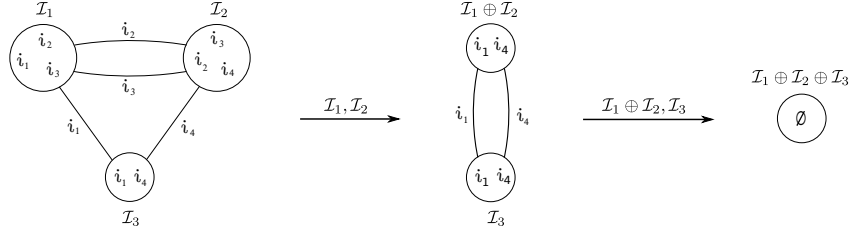
of index sets satisfying the following property:

$$\forall i \in \bigcup_{k=1}^m \mathcal{I}_k, \quad |\{j \mid i \in \mathcal{I}_j\}| = 2. \quad (2)$$

In other words, in an abstract network  $\mathcal{N}$ , each index  $i$  occurs in precisely two index sets of  $\mathcal{N}$ . We note that an index set  $\mathcal{I}$  can occur up to two times in an abstract network. We let  $|\mathcal{N}|$  denote the size of  $\mathcal{N}$ , i.e.,  $m$ . The rank of  $\mathcal{N}$ , denoted by  $\text{rank}(\mathcal{N})$ , is defined as the size of the largest index set in  $\mathcal{N}$ .

$$\text{rank}(\mathcal{N}) = \max_{j \in \{1, \dots, m\}} |\mathcal{I}_j|. \quad (3)$$

An abstract network  $\mathcal{N}$  can be intuitively visualized as a graph  $G(\mathcal{N})$  which has one vertex  $v_{\mathcal{I}}$  for each index set  $\mathcal{I} \in \mathcal{N}$ , and one edge  $e$  with endpoints  $\{\mathcal{I}, \mathcal{I}'\}$  and label  $i$ , for each pair of index sets  $\mathcal{I}, \mathcal{I}'$  with  $\mathcal{I} \cap \mathcal{I}' \neq \emptyset$  and each index  $i \in \mathcal{I} \cap \mathcal{I}'$  (Fig. 1). Note that our notion of graph of an abstract network admits multiple edges, but no loops. We say that an abstract network  $\mathcal{N}$  is connected if the graph  $G(\mathcal{N})$  associated with  $\mathcal{N}$  is connected. In this work we will only be concerned with connected abstract networks.



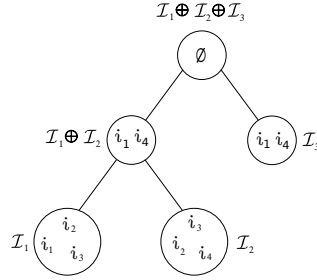
**Fig. 1** Left: the graph  $G(\mathcal{N})$  of an abstract network  $\mathcal{N} = [\mathcal{I}_1, \mathcal{I}_2, \mathcal{I}_3]$ . Middle: contracting the index sets  $\mathcal{I}_1$  and  $\mathcal{I}_2$  yields the abstract network  $\mathcal{N} = [\mathcal{I}_3, \mathcal{I}_1 \oplus \mathcal{I}_2]$ . Right: after all pairs have been contracted, the only remaining index set is the empty index set.

There is a very simple notion of contraction for abstract networks. Abstract network contractions will be used to formalize both the well known notion of tensor network contraction (Section 5), and the notion of feasibility tensor network contraction, which will be introduced in Section 6. We say that a pair of index sets  $\mathcal{I}, \mathcal{I}'$  of an abstract network  $\mathcal{N}$  is *contractible* if  $\mathcal{I} \cap \mathcal{I}' \neq \emptyset$ . In this case the contraction of  $\mathcal{I}, \mathcal{I}'$  yields the abstract network

$$\mathcal{N}' = \mathcal{N} \setminus \{\mathcal{I}, \mathcal{I}'\} \cup \{\mathcal{I} \oplus \mathcal{I}'\}$$

where  $\mathcal{I} \oplus \mathcal{I}' = \mathcal{I} \cup \mathcal{I}' \setminus (\mathcal{I} \cap \mathcal{I}')$  is the symmetric difference of  $\mathcal{I}$  and  $\mathcal{I}'$ . The contraction of a pair of index sets in an abstract network  $\mathcal{N}$  may be visualized as an operation that merges the vertices  $v_{\mathcal{I}}$  and  $v_{\mathcal{I}'}$  in the graph  $G(\mathcal{N})$  associated with  $\mathcal{N}$  (Fig. 1). Observe that in a connected abstract network with at least two vertices, there is at least one pair of contractible index sets. Additionally, when contracting a pair  $\mathcal{I}, \mathcal{I}'$  of index sets, Equation 2 ensures that the index set  $\mathcal{I} \oplus \mathcal{I}'$  is not in  $\mathcal{N}$ . Thus we have that  $|\mathcal{N}'| = |\mathcal{N}| - 1$ . Starting with an abstract network  $\mathcal{N}$  we can successively contract pairs of index-sets until we reach an abstract network whose unique index set is the empty set  $\emptyset$ . In graph-theoretic terms, starting from  $G(\mathcal{N})$  we can successively merge pairs of adjacent vertices until we reach the graph  $G([\emptyset])$  with a single vertex  $v_{\emptyset}$  (Fig. 1). Below we define the notion of contraction tree, which will be used to address both the problem of simulating an initialized quantum circuit, and the problem of computing the maximum acceptance probability of an uninitialized quantum circuit. If  $T$  is a tree, we denote by  $\text{leaves}(T)$  the set of leaves of  $T$ . We say that a node  $u \in \text{nodes}(T) \setminus \text{leaves}(T)$  is an *internal node* of  $T$ .





**Fig. 2** A contraction tree of rank 3 of the network  $\mathcal{N} = \{I_1, I_2, I_3\}$  of Fig. 1.

**Definition 3 (Contraction Tree)** Let  $\mathcal{N} = [I_1, \dots, I_m]$  be an abstract network. A *contraction tree* for  $\mathcal{N}$  is a pair  $(T, \iota)$  where  $T$  is a binary tree and  $\iota : \text{nodes}(T) \rightarrow 2^{\mathbb{N}}$  is a function that associates with each node  $u \in \text{nodes}(T)$ , an index set  $\iota(u)$  such that the following conditions are satisfied.

- (i)  $\text{leaves}(T) = \{u_1, \dots, u_m\}$  and for each  $j \in \{1, \dots, m\}$ ,  $\iota(u_j) = I_j$ .
- (ii) For each internal node  $u$ ,  $\iota(u.l) \cap \iota(u.r) \neq \emptyset$  and  $\iota(u) = \iota(u.l) \oplus \iota(u.r)$ .

Intuitively, Condition (i) says that the restriction of  $\iota$  to  $\text{leaves}(T)$  is a bijection from  $\text{leaves}(T)$  to the index sets occurring in  $\mathcal{N}$ , while Condition (ii) says that it is always possible to contract the index sets labeling the children of each internal node of  $T$ . Note that the root of  $T$  is always labeled with the empty index set  $\emptyset$ . The rank of  $(T, \iota)$  is the size of the largest index set labeling a node of  $T$ .

$$\text{rank}(T, \iota) = \max_{u \in \text{nodes}(T)} |\iota(u)|.$$

#### 4 Contraction Trees of Constant Rank and Logarithmic Height

In this section we will show that if  $\mathcal{N}$  is an abstract network whose graph  $G(\mathcal{N})$  has treewidth  $t$  and maximum degree  $\Delta$ , then one can efficiently construct a contraction tree for  $\mathcal{N}$  of rank  $O(\Delta \cdot t)$  and height  $O(\Delta \cdot t \cdot \log |\mathcal{N}|)$ . More precisely, we will prove the following theorem.

**Theorem 4 (Good Contraction Tree)** *Let  $\mathcal{N}$  be an abstract network such that the graph  $G(\mathcal{N})$  has treewidth  $t$  and maximum degree  $\Delta$ . Then one can construct in time  $2^{O(t)} \cdot |\mathcal{N}|^{O(1)}$  a contraction tree  $(T, \iota)$  for  $\mathcal{N}$  of rank  $O(\Delta \cdot t)$  and height  $O(\Delta \cdot t \cdot \log |\mathcal{N}|)$ .*

Below we define the notion of rooted carving decomposition, a variant of the notion of carving decomposition introduced by Robertson and Seymour in [23].

**Definition 4 (Rooted Carving Decomposition)** A *rooted carving decomposition* of a graph  $G = (V, E)$  is a pair  $(T, \gamma)$  where  $T = (N, F)$  is a rooted binary tree, and  $\gamma : \text{leaves}(T) \rightarrow V$  is a bijection mapping each leaf  $u \in \text{leaves}(T)$  to a single vertex  $\gamma(u) \in V$ .

Observe that the internal nodes of a carving decomposition  $(T, \gamma)$  are unlabeled. We denote by  $T[u]$  the set of nodes of the subtree of  $T$  rooted at  $u$ . Given a node  $u \in \text{nodes}(T)$  we let

$$V[u] = \gamma(\text{leaves}(T[u])) = \{v \in V \mid \exists u \in \text{leaves}(T[u]), \gamma(u) = v\} \quad (4)$$

be the set of vertices of  $G$  that are associated with some leaf in the subtree of  $T$  rooted at  $u$ . For subsets of vertices  $V_1, V_2 \subseteq V$ , let  $E(V_1, V_2)$  denote the set of edges in  $G$  with one endpoint in  $V_1$  and another endpoint in  $V_2$ . The width of  $(T, \gamma)$ , denoted  $\text{carw}(T, \gamma)$ , is defined as

$$\text{carw}(T, \gamma) = \max_{u \in \text{nodes}(T)} |E(V[u], V \setminus V[u])|. \quad (5)$$

The carving width of a graph  $G$ , denoted  $\text{carw}(G)$ , is the minimum width of a carving decomposition of  $G$ .

Next, we establish some connections between tree-decompositions and carving decompositions. Let  $G$  be a graph of treewidth  $t$ . Using the results in [22, 10], one can construct a tree decomposition  $(T, \beta)$  of  $G$  of width  $O(t)$  in time  $2^{O(t)} \cdot |G|^{O(1)}$ . From such a tree-decomposition  $(T, \beta)$  one can construct in time  $|T|^{O(1)}$  another tree-decomposition  $(T', \beta')$  of  $G$  of width  $O(t)$  and height  $O(\log |G|)$  [9]. Finally, from  $(T', \beta')$  one can construct in time  $|T'|^{O(1)}$  a carving decomposition of  $G$  of width  $O(\Delta \cdot t)$  and height  $O(\log |G|)$  [24] where  $\Delta$  is the maximum degree of  $G$ . We formalize the series of conversions we have just described into the following lemma.

**Lemma 1 ([22, 9, 24])** *Let  $G$  be a graph of maximum degree  $\Delta$  and treewidth  $t$ . One can construct in time  $2^{O(t)} \cdot |G|^{O(1)}$  a rooted carving decomposition of  $G$  of width  $O(\Delta \cdot t)$ , and height  $O(\log |G|)$ .*

For the purposes of this work we need a more well behaved notion of carving decomposition, which we call *contractive carving decomposition*.

**Definition 5 (Contractive Carving Decomposition)** We say that a rooted carving decomposition  $(T, \gamma)$  of a graph  $G = (V, E)$  is *contractive* if for each internal node  $u$  of  $T$ ,

$$E(V[u.l], V[u.r]) \neq \emptyset.$$

In other words, a rooted carving decomposition is *contractive* if for each internal node  $u$  of  $T$  there is at least one edge  $e$  of  $G$  such that one endpoint of  $e$  labels a leaf of  $T[u.l]$  and the other endpoint of  $e$  labels a leaf of  $T[u.r]$ . The next lemma states that any rooted carving decomposition of width  $w$  and height  $h$  can be transformed into a *contractive* carving decomposition of width  $w$  and height  $w \cdot h$ .

**Lemma 2** *Let  $G$  be a connected graph and  $(T, \gamma)$  be a carving decomposition of  $G$  of width  $w$  and height  $h$ . Then one can construct in time  $O(w \cdot |T|)$  a contractive carving decomposition of  $G$  of width  $w$  and height  $w \cdot h$ .*

We will prove Lemma 2 in Subsection 4.1. Before that, we will use this lemma to prove Theorem 4.

**Proof of Theorem 4.** Let  $\mathcal{N}$  be an abstract network such that  $G(\mathcal{N}) = (V, E)$  has treewidth  $t$  and maximum degree  $\Delta$ . Note that  $|G(\mathcal{N})| = |\mathcal{N}|$ . By Lemma 1, we can construct a rooted carving decomposition of  $G(\mathcal{N})$  of width  $O(\Delta \cdot t)$  and height  $O(\log |\mathcal{N}|)$ . By Lemma 2 we can convert  $(T, \gamma)$  into a contractive carving decomposition  $(T', \gamma')$  of  $G(\mathcal{N})$  of width  $O(\Delta \cdot t)$  and height  $O(\Delta \cdot t \cdot \log |\mathcal{N}|)$ .

Now, we define a function  $\iota : \text{nodes}(T') \rightarrow 2^{\mathbb{N}}$  as follows. For each leaf  $u$  of  $T'$  labeled with the vertex  $\gamma'(u) = v_{\mathcal{I}} \in G(\mathcal{N})$ , we set  $\iota(u) = \mathcal{I}$ . Therefore, at this

point we have that  $\iota$  establishes a bijection between leaves of  $T'$  and index sets of  $\mathcal{N}$ . Next, for each internal node  $u \in \text{nodes}(T')$  we set  $\iota(u) = \iota(u.l) \oplus \iota(u.r)$ . Since  $(T', \gamma')$  is a contractive carving decomposition of height  $O(\Delta \cdot t \cdot \log |\mathcal{N}|)$ , the pair  $(T', \iota)$  is a contraction tree for  $\mathcal{N}$  of height  $O(\Delta \cdot t \cdot \log |\mathcal{N}|)$ .

We claim that the rank of  $(T', \iota)$  is at most  $O(\Delta \cdot t)$ . To see this, note that for each node  $u$  of  $T'$ ,

$$\iota(u) = \bigoplus_{u' \in \text{leaves}(T'[u])} \iota(u').$$

In other words,  $\iota(u)$  is constituted by those indices that occur in precisely one leaf of the subtree  $T'[u]$  rooted at  $u$ .

$$\iota(u) = \{j \mid \text{There is a unique leaf } u' \text{ of } T'[u] \text{ such that } j \in \iota(u')\}.$$

But this implies that  $\iota(u)$  is precisely the set of indices labeling edges of  $G(\mathcal{N})$  which lie in  $E(V[u], V \setminus V[u])$ , where

$$V[u] = \{v_{\mathcal{I}} \in G(\mathcal{N}) \mid \exists u' \in \text{leaves}(T'[u]), \gamma(u') = v_{\mathcal{I}}\}.$$

Since each index  $j \in \iota(u)$  labels a unique edge in  $G(\mathcal{N})$ , we have that

$$|\iota(u)| = |E(V[u], V \setminus V[u])|.$$

Therefore, the rank of the contraction tree  $(T', \iota)$  is equal to the width  $\text{carw}(T', \gamma')$  of the carving decomposition  $(T', \gamma')$  (see Equation 5). Since, by construction,  $\text{carw}(T', \gamma') = O(\Delta \cdot t)$ , the claim follows.  $\square$

#### 4.1 Proof of Lemma 2

In this subsection we will prove Lemma 2, which states that any rooted carving decomposition  $(T, \gamma)$  of width  $w$  and height  $h$  of a graph  $G$  can be transformed into a contractive carving decomposition of width  $w$  and height  $w \cdot h$ . Recall that if  $(T, \gamma)$  is a rooted carving decomposition of a graph  $G = (V, E)$  then for each node  $u \in \text{nodes}(T)$  we let  $V[u]$  denote the set of nodes associated with the leaves in the subtree of  $T$  rooted at  $u$  (Equation 4). Below, we let  $G[u]$  denote the subgraph of  $G$  induced by the vertices in  $V[u]$ .

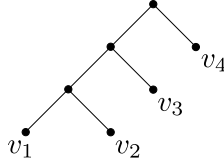
**Proposition 1** *Let  $G$  be a connected graph and  $(T, \gamma)$  be a rooted carving decomposition of  $G$  of width  $w$ . For each node  $u$  of  $T$ , the graph  $G[u]$  has at most  $w$  connected components.*

*Proof* Let  $u$  be the root of  $T$ . Since  $G$  is connected,  $G[u]$  has a unique connected component, which is  $G$  itself. Now let  $u \in \text{nodes}(T)$  be a node which is not the root of  $T$  and assume for contradiction that the connected components of  $G[u]$  are  $G_1, G_2, \dots, G_k$  for some  $k > w$ . Since  $G$  is connected, and since there are no edges between distinct connected components  $G_i$  and  $G_j$  of  $G[u]$ , we have that for each  $i \in \{1, \dots, k\}$ , there is at least one edge between a vertex of  $G_i$  and a vertex in  $V \setminus V[u]$ . But by Equation 5, this implies that the width of  $(T, \gamma)$  is at least  $k$ , contradicting in this way the assumption that the width of  $(T, \gamma)$  is  $w$ .  $\square$

Let  $G$  be a connected graph with  $n$  vertices. We say that an ordering  $v_1 v_2 \dots v_n$  of the vertices of  $G$  is a breadth first traversal in  $G$  if for every  $i, j \in \{1, \dots, n\}$  with  $i < j$ , we have that the distance from  $v_1$  to  $v_i$  is at most the distance from  $v_1$  to  $v_j$ . We note that if  $v_1 v_2 \dots v_n$  is a breadth first traversal, then for each  $k \in \{2, \dots, n\}$ , there is an edge connecting  $v_k$  to some vertex in  $\{v_1, \dots, v_{k-1}\}$ .

**Proposition 2** *Let  $G = (V, E)$  be a connected graph with  $n$  vertices and  $w$  edges. There is a contractive carving decomposition of  $G$  of height  $n - 1$  and width at most  $w$ .*

*Proof* Since  $G$  has  $w$  edges, any rooted carving decomposition of  $G$  has width at most  $w$ . Thus we just need to show that some rooted carving decomposition  $(T, \gamma)$  of  $G$  is contractive. We let  $T$  be the unique binary tree with  $n$  leaves and height  $n - 1$ . In other words,  $T$  has  $n - 1$  internal nodes, and each of these nodes has a child that is a leaf. Additionally, if  $u$  is the internal node of  $T$  farthest away from the root then both children of  $u$  are leaves. Now we define the function  $\gamma$  which is a bijection from the leaves of  $T$  to the vertices of  $G$ . Let  $v_1 v_2 \dots v_n$  be a breadth first traversal of the vertices of  $G$ . Then for each  $k \in \{2, \dots, n\}$ , there is an edge connecting  $v_k$  to some vertex in  $\{v_1, \dots, v_{k-1}\}$ . Let  $u^*$  be one of the two leaves of  $T$  at distance  $n - 1$  from the root. We set  $\gamma(u^*) = v_1$ . Now for each leaf  $u \neq u^*$  we set  $\gamma(u) = v_k$  if and only if the distance from  $u$  to the root is equal to  $n - k + 1$  (see Fig. 3).



**Fig. 3** The contractive carving decomposition corresponding to a breadth first traversal  $v_1 v_2 v_3 v_4$  of a graph with 4 vertices.

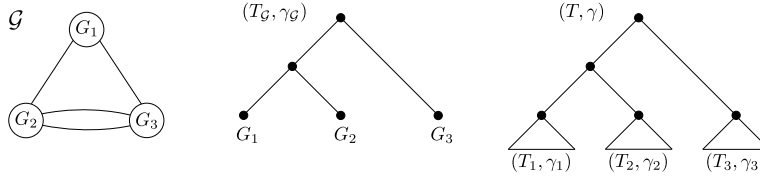
We claim that  $(T, \gamma)$  is contractive. Let  $u$  be an internal node of  $T$ . If both children of  $u$  are leaves, then both of them are at distance  $n - 1$  from the root. By definition, one of these leaves is  $u^*$ , which is labeled with  $v_1$ . The other leaf is labeled with  $v_2$ . Since  $v_1 v_2 \dots v_n$  is a breadth first traversal of the vertices of  $G$ , there is an edge connecting  $v_1$  and  $v_2$ . Now let  $u$  be an internal node for which both children are at distance  $n - k + 1$  from the root for  $k > 2$ . Then one of the children of  $u$ , say  $u.r$ , is labeled with  $v_k$  and the other leaf, say  $u.l$  is such that  $V[u.l] = \{v_1, \dots, v_{k-1}\}$ . Again, since the sequence  $v_1 v_2 \dots v_n$  is a breadth first traversal of  $G$ , we have that there is at least one edge from  $V[u.r] = \{v_k\}$  to  $V[u.l]$ . This shows that  $(T, \gamma)$  is contractive.  $\square$

Let  $G = (V, E)$  be a connected graph and let  $G_1, \dots, G_k$  be induced subgraphs of  $G$  such that  $G_i = (V_i, E_i)$  for  $i \in \{1, \dots, k\}$ ,  $V_i \cap V_j = \emptyset$  for  $i \neq j$ , and  $V = \bigcup_i V_i$ . We denote by  $\mathcal{G}(G, G_1, \dots, G_k)$  the graph with vertex set  $\mathcal{V} = \{G_1, \dots, G_k\}$ , and whose edge set  $\mathcal{E}$  has one edge  $(G_i, G_j)$  for each edge of  $G$  with one endpoint in some vertex of  $G_i$  and another endpoint in some vertex of  $G_j$ . We note that there may be multiple edges between two induced subgraphs  $G_i$  and  $G_j$ . Let  $(T_{\mathcal{G}}, \gamma_{\mathcal{G}})$  be a contractive carving decomposition of  $\mathcal{G}$  of width at most  $w$  and, for each

$i \in \{1, \dots, k\}$ , let  $(T_i, \gamma_i)$  be a contractive carving decomposition of width at most  $w$  of  $G_i$ . We denote by

$$(T, \gamma) = (T_{\mathcal{G}}, \gamma_{\mathcal{G}}) \oplus [(T_1, \gamma_1), \dots, (T_k, \gamma_k)] \quad (6)$$

the carving decomposition of  $G$  that is obtained by identifying the root of each  $(T_i, \gamma_i)$  with the leaf  $u$  of  $T_{\mathcal{G}}$  for which  $\gamma_{\mathcal{G}}(u) = G_i$  (see Fig. 4). It is immediate to check that  $(T, \gamma)$  is a contractive carving decomposition of  $G$  of width at most  $w$ .



**Fig. 4** Left: The graph  $\mathcal{G} = \mathcal{G}(G, G_1, G_2, G_3)$ , whose vertices  $G_1, G_2, G_3$  are induced subgraphs of a graph  $G$ . Middle:  $(T_{\mathcal{G}}, \gamma_{\mathcal{G}})$  is a contractive carving decomposition of  $\mathcal{G} = (G, G_1, G_2, G_3)$ . Right: For each  $i$ ,  $(T_i, \gamma_i)$  is a contractive carving decomposition of  $G_i$ .  $(T, \gamma)$  is a contractive carving decomposition of  $G$  obtained by identifying, for each  $i$ , the root of  $(T_i, \gamma_i)$  with the leaf of  $(T_{\mathcal{G}}, \gamma_{\mathcal{G}})$  labeled with  $G_i$ . If  $(T_{\mathcal{G}}, \gamma_{\mathcal{G}})$  and  $(T_i, \gamma_i)$  have width at most  $w$ , then  $(T, \gamma)$  has width at most  $w$ .

**Observation 1** *Let  $G$  be a connected graph with  $w$  edges. Then  $G$  has at most  $w + 1$  vertices.*

*Proof* The proof is by induction on the number of edges. In the base case,  $G$  has a unique edge, and therefore the observation holds trivially. Now assume that, for each  $w \geq 2$ , the observation holds for every graph with  $w - 1$  edges, and let  $G$  be a connected graph with  $w$  edges. Every such graph  $G$  can be obtained by adding an edge  $e$  to a connected graph  $G'$  with  $w - 1$  edges. By the induction hypothesis,  $G'$  has at most  $w$  vertices. Since  $G$  is connected, at least one of the vertices in  $e$  belongs to  $G'$ . Therefore  $G$  has at most  $w + 1$  vertices.

**Proof of Lemma 2** Let  $G = (V, E)$  be a connected graph, and let  $(T, \gamma)$  be a rooted carving decomposition of  $G$  of width  $w$  and height  $h$ . We will construct a *contractive* carving decomposition  $(T', \gamma')$  of  $G$  of width at most  $w$  and depth at most  $w \cdot h$ . Let  $u$  be a node of  $T$ . By proposition 1, the graph  $G[u]$  has at most  $w$  connected components. Let  $G_{u,1}, \dots, G_{u,r}$  for  $r \leq w$  be the connected components of  $G[u]$ . Let  $\text{height}(u)$  denote the height of node  $u$  in  $T$ .

*Claim* For each node  $u$  of  $T$  and each connected component  $G_{u,i}$  of  $G[u]$ , there exists a *contractive* carving decomposition  $(T_{u,i}, \gamma_{u,i})$  of  $G_{u,i}$  such that the height of  $(T_{u,i}, \gamma_{u,i})$  is at most  $w \cdot \text{height}(u)$ .

We note that Claim 4.1 implies Lemma 2, since if  $u$  is the root of  $T$  and  $G$  is connected, then the graph  $G[u]$  has a single connected component, which is  $G$  itself. The proof of Claim 4.1 is by induction on the height of the node  $u$  in  $T$ . In the base case,  $u$  is a leaf of  $T$ . In this case  $G[u]$  consists of a unique connected component  $G_{u,1}$  which is the vertex  $\gamma(u)$  of  $V$  labeling  $u$ . And therefore the carving decomposition  $(T_{u,1}, \gamma_{u,1})$  consists of a unique node  $u'$  labeled with  $\gamma_{u,1}(u') = \gamma(u)$ .

Now assume that Claim 4.1 is true for every node of  $T$  of height at most  $h$ , and let  $u$  be a node of height  $h + 1$ . By the induction hypothesis, each connected component  $G_{u.l,i}$  of  $G[u.l]$  has a contractive carving decomposition  $(T_{u.l,i}, \gamma_{u.l,i})$  of height at most  $w \cdot \text{height}(u.l)$ . Analogously, each connected component  $G_{u.r,j}$  of  $G[u.r]$  has a contractive carving decomposition  $(T_{u.r,j}, \gamma_{u.r,j})$  of height at most  $w \cdot \text{height}(u.l)$ .

Let  $G_{u,j}$  be a connected component of  $G[u]$ . Then the set of vertices of  $G_{u,j}$  is the union of the vertex sets of some connected components of  $G[u.l]$  (say  $G_{u.l,1}, \dots, G_{u.l,p}$ ), and some connected components of  $G[u.r]$  (say  $G_{u.r,1}, \dots, G_{u.r,q}$ ). Note that all edges of  $G_{u,j}$  that do not belong to some of these components, must connect some vertex of  $G_{u.l,i}$  to some vertex of  $G_{u.r,i'}$  for some  $i \in \{1, \dots, p\}$  and some  $i' \in \{1, \dots, q\}$ . But since the carving decomposition  $(T, \gamma)$  of the graph  $G$  has width at most  $w$ , there can be at most  $w$  such new edges. In other words, the graph  $\mathcal{G} = \mathcal{G}(G_{u,j}, G_{u.l,1}, \dots, G_{u.l,p}, G_{u.r,1}, \dots, G_{u.r,q})$  has at most  $w$  edges. Since  $\mathcal{G}$  is connected, by Observation 1 we have that  $\mathcal{G}$  has at most  $w + 1$  vertices. By Proposition 2,  $\mathcal{G}$  has a contractive carving decomposition  $(T_{\mathcal{G}}, \gamma_{\mathcal{G}})$  of height at most  $w$  and width at most  $w$ . Therefore the carving decomposition

$$(T_{u,i}) = (T_{\mathcal{G}}, \gamma_{\mathcal{G}}) \oplus [(T_{u.l,1}, \gamma_{u.l,1}), \dots, (T_{u.l,p}, \gamma_{u.l,p}), (T_{u.r,1}, \gamma_{u.r,1}), \dots, (T_{u.r,q}, \gamma_{u.r,q})]$$

is contractive and has width at most  $w$ . Since  $(T_{\mathcal{G}}, \gamma_{\mathcal{G}})$  has height at most  $w$ , and by assumption, each  $(T_{u.l,i}, \gamma_{u.l,i})$  and  $(T_{u.r,j}, \gamma_{u.r,j})$  has height at most  $w \cdot h$ , we have that the height of  $(T_{u,i}, \gamma_{u,i})$  is at most  $w \cdot h + w = w \cdot (h + 1)$ . This proves Claim 4.1, and therefore also Lemma 2, by letting  $u$  be the root of  $T$ .  $\square$

## 5 Tensor Networks

In this section we will redefine the well known notion of tensor network in function of abstract networks. Within this formalism, a tensor network is a pair  $(\mathcal{N}, \lambda)$  where  $\mathcal{N}$  is an abstract network, and  $\lambda$  is a function that associates a tensor  $\lambda(\mathcal{I})$  of rank  $|\mathcal{I}|$  with each index set  $\mathcal{I} \in \mathcal{N}$ . We believe that defining tensor networks in this way has the advantage of separating the algorithmic aspects of tensor networks from their quantum aspects. Additionally, the formalism of abstract networks will also be used in Section 6 to introduce the notion of *feasibility tensor networks* which will be used to address the problem of approximating the maximum acceptance probability of quantum circuits with uninitialized inputs.

Let  $\Pi(d) = \{|b_1\rangle\langle b_2| \mid b_1, b_2 \in \{0, \dots, d - 1\}\}$ . A  $d$ -state tensor with index set  $\mathcal{I} = \{i_1, \dots, i_k\}$  is an array  $g$  consisting of  $|\Pi(d)|^k = d^{2k}$  complex numbers. The entries

$$g(\sigma_{i_1}, \dots, \sigma_{i_k})$$

of  $g$  are indexed by a sequence of variables  $\sigma_{i_1}, \dots, \sigma_{i_k}$ , each of which ranges over the set  $\Pi(d)$ . We note that if  $\mathcal{I} = \emptyset$  then a tensor with index set  $\mathcal{I}$  is simply a complex number  $g(-)$ . If  $g$  is a tensor with index set  $\mathcal{I}$  then we let  $\text{rank}(g) = |\mathcal{I}|$  be the rank of  $g$ . We denote by  $\mathbb{T}(d, \mathcal{I})$  the set of all  $d$ -state tensors with index set  $\mathcal{I}$  and by  $\mathbb{T}(d) = \bigcup_{\mathcal{I} \subseteq \mathbb{N}} \mathbb{T}(d, \mathcal{I})$  the set of all  $d$ -state tensors.

**Definition 6 (Tensor Network)** A *tensor network* is a pair  $(\mathcal{N}, \lambda)$  where  $\mathcal{N}$  is an abstract network and  $\lambda$  is a function that associates with each index set  $\mathcal{I} \in \mathcal{N}$ , a tensor  $\lambda(\mathcal{I}) \in \mathbb{T}(d, \mathcal{I})$ .

A tensor network  $(\mathcal{N}, \lambda)$  is connected if  $\mathcal{N}$  is connected. In this work we will only be interested in connected tensor networks. An important operation involving tensors is the operation of tensor contraction. If  $g$  is a tensor with index set  $\mathcal{I} = \{i_1, \dots, i_k, l_1, \dots, l_r\}$  and  $g'$  is a tensor with index set  $\mathcal{I}' = \{j_1, \dots, j_{k'}, l_1, \dots, l_r\}$  then the contraction of  $g$  and  $g'$  gives rise to the tensor  $\text{Contr}(g, g')$  with index set  $\mathcal{I} \oplus \mathcal{I}' = \{i_1, \dots, i_k, j_1, \dots, j_{k'}\}$  where each entry  $\text{Contr}(g, g')(\sigma_{i_1}, \dots, \sigma_{i_k}, \sigma_{j_1}, \dots, \sigma_{j_{k'}})$  is defined as

$$\sum_{\sigma_{l_1}, \dots, \sigma_{l_r} \in \Pi(d)} g(\sigma_{i_1}, \dots, \sigma_{i_k}, \sigma_{l_1}, \dots, \sigma_{l_r}) \cdot g'(\sigma_{j_1}, \dots, \sigma_{j_{k'}}, \sigma_{l_1}, \dots, \sigma_{l_r}). \quad (7)$$

If  $(\mathcal{N}, \lambda)$  is a tensor network and  $\mathcal{I}_1, \mathcal{I}_2$  is a pair of contractible sets in  $\mathcal{N}$  then we say that the tensor network  $(\mathcal{N}', \lambda')$  is obtained from  $(\mathcal{N}, \lambda)$  by the contraction of  $\mathcal{I}_1$  and  $\mathcal{I}_2$  if  $\mathcal{N}' = (\mathcal{N} \setminus \{\mathcal{I}_1, \mathcal{I}_2\}) \cup \{\mathcal{I}_1 \oplus \mathcal{I}_2\}$ , and if  $\lambda'$  satisfies the following conditions.

1.  $\lambda'(\mathcal{I}_1 \oplus \mathcal{I}_2) = \text{Contr}(\lambda(\mathcal{I}_1), \lambda(\mathcal{I}_2))$ .
2.  $\lambda'(\mathcal{I}) = \lambda(\mathcal{I})$  for each  $\mathcal{I} \in \mathcal{N}' \setminus \{\mathcal{I}_1 \oplus \mathcal{I}_2\}$ .

Any connected tensor network with  $m$  index sets can be contracted  $m-1$  times. The result of this contraction process is a tensor network  $([\emptyset], \lambda_0)$  with a unique index set, namely  $\emptyset$ , which is labeled with a rank-0 tensor  $\lambda_0(\emptyset)$  (that is to say, a complex number). The value of  $(\mathcal{N}, \lambda)$ , denoted by  $\text{val}(\mathcal{N}, \lambda)$ , is defined as the absolute value of  $\lambda_0(\emptyset)$ . More precisely,  $\text{val}(\mathcal{N}, \lambda) = |\lambda_0(\emptyset)|$ . We observe that the value of a tensor network is well defined, since it does not depend on the order in which the tensors of the network are contracted.

### 5.1 Mapping Quantum Circuits with Initialized Inputs to Tensor Networks

One of the main reasons behind the popularity of tensor networks is the fact that they can be used to simulate quantum circuits. First, we note that both density operators and quantum gates can be naturally regarded as tensors. If  $\rho$  is a density operator acting on  $d$ -dimensional qudits indexed by  $\mathcal{I} = \{i_1, \dots, i_k\}$ , then the tensor  $\rho$  associated with  $\rho$  is defined as

$$\rho(\sigma_{i_1}, \dots, \sigma_{i_k}) = \text{tr} \left( \rho \cdot [\sigma_{i_1}^\dagger \otimes \dots \otimes \sigma_{i_k}^\dagger] \right). \quad (8)$$

If  $Q$  is a quantum gate with inputs indexed by  $\mathcal{I} = \{i_1, \dots, i_k\}$  and outputs indexed by  $\mathcal{I}' = \{j_1, \dots, j_l\}$  where  $\mathcal{I} \cap \mathcal{I}' = \emptyset$ , then the tensor  $Q$  associated with  $Q$  is defined as

$$Q(\sigma_{i_1}, \dots, \sigma_{i_k}, \sigma_{j_1}, \dots, \sigma_{j_l}) = \text{tr} \left( Q \cdot [\sigma_{i_1} \otimes \dots \otimes \sigma_{i_k}] \cdot [\sigma_{j_1}^\dagger \otimes \dots \otimes \sigma_{j_l}^\dagger] \right). \quad (9)$$

In the sequel, we will not distinguish between gates or density matrices and their associated tensors. If  $C = (V, E, \theta, \xi)$  is a quantum circuit in which all inputs are initialized, then the tensor network  $(\mathcal{N}_C, \lambda_C)$  associated with  $C$  is obtained as follows. For each vertex  $v \in V$ , let  $\mathcal{I}(v)$  be the index set consisting of all integers labeling edges of  $C$  which are incident with  $v$ . Then we add  $\mathcal{I}(v)$  to  $\mathcal{N}_C$  and set  $\lambda_C(\mathcal{I}(v))$  to be the tensor associated with the gate  $\theta(v)$  of  $C$ . We say that  $\mathcal{N}_C$  is the abstract network associated with  $C$ . The following proposition, which is well known

in tensor-network theory (see [20] for a proof), establishes a close correspondence between the value of tensor networks and the acceptance probability of quantum circuits.

**Proposition 3** *Let  $C$  be a quantum circuit with  $n$  inputs initialized with the state  $|y\rangle$  for some  $y \in \{0, \dots, d-1\}^n$ . Then  $\text{val}(\mathcal{N}_C, \lambda_C) = \text{Pr}(C, |y\rangle)$ .*

In other words,  $\text{val}(\mathcal{N}_C, \lambda_C)$  is the acceptance probability of  $C$ .

## 5.2 Computing the Value of a Tensor Network

The process of computing the value  $\text{val}(\mathcal{N}, \lambda)$  of a tensor network  $(\mathcal{N}, \lambda)$  is known as simulation. Given a contraction tree  $(T, \iota)$  of rank  $r$  for  $\mathcal{N}$ , the following definition can be used to compute  $\text{val}(\mathcal{N}, \lambda)$  in time  $d^{O(r)} \cdot |\mathcal{N}|^{O(1)}$ .

**Definition 7 (Tensor Network Simulation)** Let  $(\mathcal{N}, \lambda)$  be a tensor network and  $(T, \iota)$  be a contraction tree for  $\mathcal{N}$ . A simulation of  $(\mathcal{N}, \lambda)$  on  $(T, \iota)$  is a function  $\hat{\lambda} : \text{nodes}(T) \rightarrow \mathbb{T}(d)$  satisfying the following conditions:

1. For each leaf  $u$  of  $T$ ,  $\hat{\lambda}(u) = \lambda(\iota(u))$ .
2. For each internal node  $u$  of  $T$ ,  $\hat{\lambda}(u) = \text{Contr}(\hat{\lambda}(u.l), \hat{\lambda}(u.r))$ .

Note that if  $u$  is the root of a contraction tree, then  $\iota(u) = \emptyset$ . In this case the tensor  $\hat{\lambda}(u)$  is a rank-0 tensor (that is, a complex number) which is obtained by contracting all tensors in  $(\mathcal{N}, \lambda)$ . This implies that  $|\hat{\lambda}(u)| = \text{val}(\mathcal{N}, \lambda)$ . If the contraction tree  $(T, \iota)$  has rank  $r$ , then for each node  $u$  of  $T$  we have that  $|\iota(u)| \leq r$ . In other words, for each  $u \in \text{nodes}(T)$ , the tensor  $\hat{\lambda}(u)$  has rank at most  $r$ , and for this reason  $\hat{\lambda}(u)$  can be represented by  $d^{2r}$  complex numbers. In this way, the simulation  $\hat{\lambda}$  can be inductively constructed in time  $d^{O(r)} \cdot |\mathcal{N}|^{O(1)}$ , and therefore,  $\text{val}(\mathcal{N}, \lambda)$  can be computed in time  $d^{O(r)} \cdot |\mathcal{N}|^{O(1)}$ .

By Theorem 4, if the graph  $G(\mathcal{N})$  of an abstract network  $\mathcal{N}$  has treewidth  $t$  and maximum degree  $\Delta$ , then one can construct in polynomial time a contraction tree for  $\mathcal{N}$  of rank  $O(\Delta \cdot t)$ . Therefore, Definition 7 can be used to compute the value  $\text{val}(\mathcal{N}, \lambda)$  of a tensor network  $(\mathcal{N}, \lambda)$  in time  $d^{O(\Delta \cdot t)} \cdot |\mathcal{N}|^{O(1)}$ . Now let  $\mathcal{U}$  be a fixed finite universal set of gates, and let  $C$  be a quantum circuit of treewidth  $t$ , whose inputs are initialized with a basis state  $|y\rangle$ , and whose gates are drawn from  $\mathcal{U}$ . Let  $(\mathcal{N}_C, \lambda_C)$  be the tensor network associated with  $C$ . By Proposition 3,  $\text{val}(\mathcal{N}_C, \lambda_C) = \text{Pr}(C, |y\rangle)$ . Therefore, the simulation algorithm described above can be used to compute the acceptance probability of  $C$  in time  $d^{O(\Delta(\mathcal{U}) \cdot t)} \cdot |C|^{O(1)}$ , where  $|C|$  is the number of vertices of  $C$ . We note that this algorithm has the same asymptotic time complexity as the original contraction algorithm for tensor networks devised in [20], although our contraction technique based on Theorem 4 is different from that employed in [20].

We observe that the fact that the contraction trees constructed in Theorem 4 have logarithmic height is not relevant for the time complexity of the simulation algorithm described above. Nevertheless, as we will see in Section 7, contraction trees of logarithmic height will be essential when devising a polynomial time algorithm for the problem of approximating the maximum acceptance probability of constant-treewidth quantum circuits with uninitialized inputs. Even though it is possible to extract contraction trees of constant rank from the contraction



sequences defined in [20], the contraction trees obtained in this way are not guaranteed to have logarithmic height. Therefore, our contraction algorithm cannot be directly replaced by the contraction algorithm devised in [20] when addressing the satisfiability of constant-treewidth quantum circuits.

## 6 Feasibility Tensor Networks

In Section 5 we defined tensor networks in terms of abstract networks and showed how contraction trees can be used to address the problem of computing the value of a tensor network. In this section we will use abstract networks to introduce *feasibility tensor networks*. We will then proceed to show that feasibility tensor networks can be used to address the problem of computing an assignment that maximizes the acceptance probability of quantum circuits with uninitialized inputs.

**Definition 8 (Feasibility Tensor Network)** A *feasibility tensor network* is a pair  $(\mathcal{N}, \Lambda)$  where  $\mathcal{N}$  is an abstract network and  $\Lambda : \mathcal{N} \rightarrow 2^{\mathbb{T}(d)}$  is a function that associates with each index set  $\mathcal{I} \in \mathcal{N}$  a finite set of tensors  $\Lambda(\mathcal{I}) \subseteq \mathbb{T}(d, \mathcal{I})$ .

Note that the only difference between tensor networks and feasibility tensor networks is that while in the former we associate a tensor with each index set, in the latter we associate a set of tensors with each index set. If  $(\mathcal{N}, \Lambda)$  is a feasibility tensor network, then an *initialization* of  $(\mathcal{N}, \Lambda)$  is a function  $\lambda : \mathcal{N} \rightarrow \mathbb{T}(d)$  such that  $\lambda(\mathcal{I}) \in \Lambda(\mathcal{I})$  for each index set  $\mathcal{I} \in \mathcal{N}$ . Intuitively, an initialization  $\lambda$  chooses one tensor  $\lambda(\mathcal{I})$  from each set of tensors  $\Lambda(\mathcal{I})$ . Observe that for each such an initialization  $\lambda$ , the pair  $(\mathcal{N}, \lambda)$  is a tensor network as defined in Section 5. The value of a feasibility tensor network is defined as

$$\text{VAL}(\mathcal{N}, \Lambda) = \max\{\text{val}(\mathcal{N}, \lambda) \mid \lambda \text{ is an initialization of } (\mathcal{N}, \Lambda)\}. \quad (10)$$

Below we show that the problem of finding an assignment that maximizes the acceptance probability of a quantum circuit with uninitialized inputs can be reduced to the problem of computing an initialization of maximum value for a feasibility tensor network. Therefore, the problem of computing the value of a feasibility tensor network is QCMA hard. The conversion from quantum circuits with uninitialized inputs to feasibility tensor networks goes as follows: Each uninitialized input  $v$  corresponds to the set of tensors  $\{|0\rangle\langle 0|, |1\rangle\langle 1|, \dots, |d-1\rangle\langle d-1|\}$ . Intuitively, this set of tensors consists of all possible values that can be used to initialize  $v$ . On the other hand, each input vertex  $v$  which is already initialized with a density matrix  $|i\rangle\langle i|$  corresponds to the singleton set  $\{|i\rangle\langle i|\}$ . Finally, each gate  $g$  of the circuit corresponds to the singleton set  $\{g\}$ . We formalize this construction in Definition 9.

**Definition 9 (From Quantum Circuits to Feasibility Tensor Networks)** Let  $C = (V, E, \theta, \xi)$  be a quantum circuit in which some of the inputs are uninitialized. The feasibility tensor network associated with  $C$  is denoted by  $(\mathcal{N}_C, \Lambda_C)$ , where  $\mathcal{N}_C = \{\mathcal{I}(v) \mid v \in V\}$  is the abstract network associated with  $C$ , and  $\Lambda_C$  is such that for each  $v \in V$ ,

$$\Lambda_C(\mathcal{I}(v)) = \begin{cases} \{|0\rangle\langle 0|, |1\rangle\langle 1|, \dots, |d-1\rangle\langle d-1|\} & \text{if } v \text{ is an uninitialized input,} \\ \{\theta(v)\} & \text{otherwise.} \end{cases} \quad (11)$$

Now let  $\lambda$  be an initialization of the feasibility tensor network  $(\mathcal{N}_C, \Lambda_C)$ . Then the tensor network  $(\mathcal{N}_C, \lambda)$  is precisely the tensor network associated with the circuit  $C$  in which the inputs are initialized with the state

$$|y_\lambda\rangle = \otimes_{v \in \text{In}(C)} \lambda(\mathcal{I}(v)). \quad (12)$$

In other words,  $\text{val}(\mathcal{N}_C, \lambda) = \text{Pr}(C, |y_\lambda\rangle)$ . Therefore, we have the following observation.

**Observation 2** *For each quantum circuit  $C$ , the value  $\text{VAL}(\mathcal{N}_C, \Lambda_C)$  of the feasibility tensor network  $(\mathcal{N}_C, \Lambda_C)$  associated with  $C$  is equal to the maximum acceptance probability  $\text{Pr}^{\text{cl}}(C)$  of  $C$ .*

## 7 Approximating the Value of a Feasibility Tensor Network

In this section we will devise an algorithm that, when given a feasibility tensor network  $(\mathcal{N}, \Lambda)$  and a real number  $\delta \in (0, 1)$  as input, can be used both to approximate the value  $\text{VAL}(\mathcal{N}, \Lambda)$  up to a  $\delta$  additive factor, and to construct an initialization  $\lambda$  such that  $|\text{VAL}(\mathcal{N}, \Lambda) - \text{val}(\mathcal{N}, \lambda)| \leq \delta$ . In particular, our algorithm runs in polynomial time if we are given a contraction tree for  $\mathcal{N}$  of constant rank and logarithmic height. As we saw in Section 4 if the graph  $G(\mathcal{N})$  associated with  $\mathcal{N}$  has constant treewidth and constant maximum degree, then a contraction tree with these properties can be efficiently constructed using Theorem 4.

### 7.1 Tensor $\varepsilon$ -Nets

We start by defining suitable notions of norm and distance for tensors. If  $g$  is a tensor with index-set  $\mathcal{I} = \{i_1, \dots, i_k\}$ , then the  $\mathcal{L}_\infty$  norm of  $g$  is defined as

$$\|g\| = \max_{\sigma_{i_1} \dots \sigma_{i_k}} |g(\sigma_{i_1}, \dots, \sigma_{i_k})|, \quad (13)$$

where for each  $j \in \{1, \dots, k\}$ ,  $\sigma_{i_j}$  ranges over the set  $\Pi(d)$ , and  $|g(\sigma_{i_1}, \dots, \sigma_{i_k})|$  is the absolute value of the entry  $g(\sigma_{i_1}, \dots, \sigma_{i_k})$  of  $g$ . Having the definition of norm of a tensor in hands, we define the distance between two tensors  $g$  and  $g'$  as  $\|g - g'\|$ . The next step consists in defining a suitable notion of  $\varepsilon$ -net of tensors.

**Definition 10 (Tensor  $(d, \varepsilon, \mathcal{I})$ -Net)** Let  $\mathcal{I}$  be an index set,  $d \in \mathbb{N}$ , and  $\varepsilon \in \mathbb{R}$  with  $0 < \varepsilon < 1$ . A tensor  $(d, \varepsilon, \mathcal{I})$ -net is a set  $\mathbb{T}(d, \varepsilon, \mathcal{I})$  of  $d$ -state tensors with index set  $\mathcal{I}$  such that for each  $g$  in  $\mathbb{T}(d, \mathcal{I})$ , there exists  $g' \in \mathbb{T}(d, \varepsilon, \mathcal{I})$  with  $\|g - g'\| \leq \varepsilon$ .

It is straightforward to construct a  $(d, \varepsilon, \mathcal{I})$ -net for tensors. We just need to consider the set of all  $d$ -state tensors with index set  $\mathcal{I}$  in which each entry is a complex number of the form  $a + b \cdot i$  for  $-1 \leq a, b \leq 1$  and  $a, b$  integer multiples of  $\varepsilon/2$ . We observe that we do not need to assume that the tensors in our  $(d, \varepsilon, \mathcal{I})$ -net correspond to physically realizable operations. Our approximation algorithm does not need this assumption. Since a  $d$ -state tensor over the index set  $\mathcal{I}$  has  $d^{2|\mathcal{I}|}$  entries, we have the following proposition upper bounding the size of a tensor  $(d, \varepsilon, \mathcal{I})$ -net.

**Proposition 4** *For each index set  $\mathcal{I}$ , each  $d \in \mathbb{N}$  and each  $\varepsilon \in \mathbb{R}$  with  $0 < \varepsilon < 1$ , one can construct a tensor  $(d, \varepsilon, \mathcal{I})$ -net  $\mathbb{T}(d, \varepsilon, \mathcal{I})$  with at most  $(1/\varepsilon)^{\exp(O(|\mathcal{I}| \log d))}$  tensors.*

If  $g$  is a tensor in  $\mathbb{T}(d, \mathcal{I})$ , then we let  $\text{Trunc}_\varepsilon(g)$  be an arbitrary tensor  $g'$  in  $\mathbb{T}(d, \varepsilon, \mathcal{I})$  such that  $\|g - g'\| \leq \varepsilon$ . Going further, if  $\mathcal{F}$  is a set of tensors then we let

$$\text{Trunc}_\varepsilon(\mathcal{F}) = \{\text{Trunc}_\varepsilon(g) \mid g \in \mathcal{F}\}$$

be the truncated version of  $\mathcal{F}$ .

## 7.2 Approximation Algorithm

In this subsection we will address the problem of  $\delta$ -approximating the value of feasibility tensor networks and the problem of finding  $\delta$ -optimal initializations for feasibility tensor networks. First, we define the notion of contraction for pairs of sets of tensors. Let  $\mathcal{I}, \mathcal{I}'$  be a pair of index sets with  $\mathcal{I} \cap \mathcal{I}' \neq \emptyset$ . Let  $\mathcal{F} \subseteq \mathbb{T}(d, \mathcal{I})$  be a finite set of tensors with index set  $\mathcal{I}$  and  $\mathcal{F}' \subseteq \mathbb{T}(d, \mathcal{I}')$  be a finite set of tensors with index set  $\mathcal{I}'$ . The contraction of  $\mathcal{F}$  and  $\mathcal{F}'$  is defined as

$$\text{Contr}(\mathcal{F}, \mathcal{F}') = \{\text{Contr}(g, g') \mid g \in \mathcal{F}, g' \in \mathcal{F}'\}. \quad (14)$$

Subsequently, we define a notion of simulation for feasibility tensor networks that is analog to our definition of simulation for tensor networks introduced in Subsection 5.2. The simulation of a feasibility tensor network  $(\mathcal{N}, \Lambda)$  on a contraction tree  $(T, \iota)$  is a function  $\hat{A}$  that associates a set of tensors with each node of  $T$ . First, with each leaf  $u$  of  $T$  such that  $\iota(u) = \mathcal{I}$ , we associate the set of tensors  $\hat{A}(\mathcal{I}) = \Lambda(\mathcal{I})$ . Then, with each internal node  $u$  of  $T$ , we associate the set of tensors  $\hat{A}(u) = \text{Trunc}_\varepsilon(\text{Contr}(\hat{A}(u.l), \hat{A}(u.r)))$ . We note that the truncation is necessary to keep the size of each set from growing exponentially as the contractions take place. This construction is given more formally in Definition 11 below.

**Definition 11 (Feasibility Tensor Network Simulation)** Let  $(\mathcal{N}, \Lambda)$  be a feasibility tensor network and  $(T, \iota)$  be a contraction tree for  $\mathcal{N}$ . An  $\varepsilon$ -simulation of  $(\mathcal{N}, \Lambda)$  on  $(T, \iota)$  is a function  $\hat{A}: N \rightarrow 2^{\mathbb{T}(d, \varepsilon)}$  satisfying the following properties:

1. For each leaf  $u$  of  $T$ ,  $\hat{A}(u) = \Lambda(\iota(u))$ ,
2. For each internal node  $u$  of  $T$ ,  $\hat{A}(u) = \text{Trunc}_\varepsilon(\text{Contr}(\hat{A}(u.l), \hat{A}(u.r)))$ .

Intuitively, an  $\varepsilon$ -simulation  $\hat{A}$  is a function that keeps track of all ways of simulating tensor networks  $(\mathcal{N}, \lambda)$  where  $\lambda$  is an arbitrary initialization of  $(\mathcal{N}, \Lambda)$ . In particular, if  $u$  is the root of  $(T, \iota)$  then  $u$  is labeled with a set  $\hat{A}(u)$  of complex numbers. For each such complex number  $a$ , there exists an initialization  $\lambda$  of  $(\mathcal{N}, \Lambda)$  such that  $|a|$  is an approximation of  $\text{val}(\mathcal{N}, \lambda)$ . Conversely, for each initialization  $\lambda$  of  $(\mathcal{N}, \Lambda)$ , there exists some number  $a \in \hat{A}(u)$  such that  $|a|$  approximates  $\text{val}(\mathcal{N}, \lambda)$ . Therefore, the maximum absolute value  $\alpha$  of a complex number in  $\hat{A}(u)$  is an approximation of  $\text{VAL}(\mathcal{N}, \Lambda)$ . An actual initialization  $\lambda$  of  $(\mathcal{N}, \Lambda)$  such that  $\text{val}(\mathcal{N}, \lambda) = \alpha \approx \text{VAL}(\mathcal{N}, \Lambda)$  can be found by backtracking. Theorem 5 below, which will be proved in Subsection 7.3, establishes an upper bound for the time complexity and for the error of the approximation scheme described above. The error of such process depends exponentially on the height of the contraction tree, while the time complexity depends exponentially on the rank of the contraction tree.

**Theorem 5 (Feasibility Tensor Network Satisfiability)** *Let  $(\mathcal{N}, A)$  be a feasibility tensor network,  $(T, \iota)$  be a contraction tree for  $\mathcal{N}$  of rank  $r$  and height  $h$ , and  $\varepsilon$  be a real number with  $0 < \varepsilon < 1$ .*

1. *One can compute a number  $\alpha$  such that  $|\alpha - \text{VAL}(\mathcal{N}, A)| \leq \varepsilon \cdot (3d^{2r} + 1)^h$  in time  $|\mathcal{N}| \cdot \varepsilon^{-\exp(O(r \log d))}$ .*
2. *One can construct in time  $|\mathcal{N}| \cdot \varepsilon^{-\exp(O(r \log d))}$  an initialization  $\lambda$  of  $(\mathcal{N}, A)$  such that*

$$|\text{val}(\mathcal{N}, \lambda) - \text{VAL}(\mathcal{N}, A)| \leq \varepsilon \cdot (3d^{2r} + 1)^h.$$

We note that to efficiently compute  $\alpha$  and  $\lambda$  in Theorem 5 above, we need to have in hands a contraction tree for  $\mathcal{N}$  whose rank is bounded by a constant, and whose height is bounded by  $O(\log |\mathcal{N}|)$ . The next theorem (Theorem 6) states that approximately optimal initializations of feasibility tensor networks of constant treewidth and constant maximum degree can be computed in polynomial time. Note that in this case, the existence of a contraction tree of constant rank and logarithmic height is guaranteed by Theorem 4.

**Theorem 6** *Let  $(\mathcal{N}, A)$  be a feasibility tensor network such that the graph  $G(\mathcal{N})$  has treewidth  $t$  and maximum degree  $\Delta$ . For each  $\delta$  with  $1/\text{poly}(n) < \delta < 1$ , one can compute in time  $(|\mathcal{N}|/\delta)^{\exp(O(\Delta \cdot t \cdot \log d))}$  an initialization  $\lambda$  of  $(\mathcal{N}, A)$  such that*

$$|\text{val}(\mathcal{N}, \lambda) - \text{VAL}(\mathcal{N}, A)| \leq \delta.$$

*Proof* By Theorem 4, we can construct a contraction tree for  $\mathcal{N}$  of rank  $r = O(\Delta \cdot t)$  and height  $h = O(\Delta \cdot t \cdot \log |\mathcal{N}|)$ . By Theorem 5.2, we can compute in time  $|\mathcal{N}|^{O(1)} \cdot \varepsilon^{-\exp(O(r \log d))}$  an initialization  $\lambda$  of  $(\mathcal{N}, A)$  such that

$$|\text{val}(\mathcal{N}, \lambda) - \text{VAL}(\mathcal{N}, A)| \leq \varepsilon \cdot 2^{O(r \cdot h \cdot \log d)}.$$

Therefore, by setting  $\varepsilon = \delta/|\mathcal{N}|^{O(\Delta^2 \cdot t^2 \cdot \log d)}$ , we can compute an initialization  $\lambda$  for  $(\mathcal{N}, A)$  such that  $|\text{val}(\mathcal{N}, \lambda) - \text{VAL}(\mathcal{N}, A)| \leq \delta$  in time  $(|\mathcal{N}|/\delta)^{\exp(O(\Delta \cdot t \cdot \log d))}$ .  $\square$

Since the problem of computing a  $\delta$ -optimal initialization of a quantum circuit can be reduced to the problem of computing a  $\delta$ -optimal initialization of a feasibility tensor network, our main theorem (Theorem 1) follows from Theorem 6.

**Proof of Theorem 1** Let  $C$  be a quantum circuit with  $n$  uninitialized inputs, treewidth  $t$ , and  $\text{poly}(n)$  gates drawn from a finite universal set of gates  $\mathcal{U}$ . Let  $(\mathcal{N}_C, A_C)$  be the feasibility tensor network associated with  $C$  according to Definition 9. Then the graph  $G(\mathcal{N}_C)$  has treewidth  $t$ , and maximum degree  $\Delta(\mathcal{U})$ , where  $\Delta(\mathcal{U})$  is the maximum number of inputs and outputs of a gate in  $\mathcal{U}$ . Additionally,  $\text{Pr}^{cl}(C) = \text{VAL}(\mathcal{N}_C, A_C)$  and, by Equation 12, each initialization  $\lambda$  of  $(\mathcal{N}_C, A_C)$  corresponds to a classical initialization  $|y_\lambda\rangle$  of  $C$  such that  $\text{Pr}^{cl}(C, |y_\lambda\rangle) = \text{val}(\mathcal{N}_C, \lambda)$ . Using Theorem 6, we can compute in time  $(n/\delta)^{\exp(O(\Delta(\mathcal{U}) \cdot t \cdot \log d))}$  an initialization  $\lambda$  of  $(\mathcal{N}_C, A_C)$  such that

$$|\text{val}(\mathcal{N}_C, \lambda) - \text{VAL}(\mathcal{N}_C, A_C)| \leq \delta.$$

Since  $\text{VAL}(\mathcal{N}_C, A_C) = \text{Pr}^{cl}(C)$ , we have  $|\text{Pr}(C, |y_\lambda\rangle) - \text{Pr}^{cl}(C)| \leq \delta$ . Finally, since  $\Delta(\mathcal{U})$  and  $d$  are constants, the time complexity of this construction can be simplified to  $(n/\delta)^{O(t)}$ .  $\square$

## 7.3 Proof of Theorem 5

In this subsection we will prove Theorem 5. In the proof we will devise an algorithm to  $\delta$ -approximate the value of a feasibility tensor network and to compute an initialization which achieves a near optimal value. We start by stating a couple of auxiliary lemmas. In particular, the following lemma establishes an upper bound for the distance between the contraction of two given tensors  $g_1$  and  $g_2$  and the contraction of approximations  $g'_1$  and  $g'_2$  of  $g_1$  and  $g_2$  respectively.

**Lemma 3** *Let  $g_1$  and  $g'_1$  be tensors with index set  $\mathcal{I}_1$  and let  $g_2$  and  $g'_2$  be tensors with index set  $\mathcal{I}_2$ , where  $\mathcal{I}_1 \cap \mathcal{I}_2 \neq \emptyset$ . Let  $\|g_1 - g'_1\| \leq \varepsilon$  and  $\|g_2 - g'_2\| \leq \varepsilon$ . Then*

$$\|Contr(g_1, g_2) - Contr(g'_1, g'_2)\| \leq \varepsilon \cdot 3d^{2 \cdot |\mathcal{I}_1 \cap \mathcal{I}_2|}.$$

*Proof* Let  $\mathcal{I}_1 = \{i_1, \dots, i_k, l_1, \dots, l_r\}$  and  $\mathcal{I}_2 = \{j_1, \dots, j_{k'}, l_1, \dots, l_r\}$  be index sets. Let  $g'_1 = g_1 + e_1$  and  $g'_2 = g_2 + e_2$  where  $e_1$  and  $e_2$  are offset tensors. Since  $\|g_1 - g'_1\| \leq \varepsilon$  and  $\|g_2 - g'_2\| \leq \varepsilon$ , we have that  $\|e_1\| \leq \varepsilon$  and  $\|e_2\| \leq \varepsilon$ . Let  $\sigma = (\sigma_{i_1}, \dots, \sigma_{i_k}, \sigma_{l_1}, \dots, \sigma_{l_r})$  and  $\sigma' = (\sigma_{j_1}, \dots, \sigma_{j_{k'}}, \sigma_{l_1}, \dots, \sigma_{l_r})$ . Then we have that

$$Contr(g'_1, g'_2)(\sigma_{i_1}, \dots, \sigma_{i_k}, \sigma_{j_1}, \dots, \sigma_{j_{k'}}) = \sum_{\sigma_{l_1}, \dots, \sigma_{l_r} \in \Pi(d)} (g_1(\sigma) + e_1(\sigma))(g_2(\sigma') + e_2(\sigma')).$$

By reorganizing the right hand side,  $Contr(g'_1, g'_2)(\sigma_{i_1}, \dots, \sigma_{i_k}, \sigma_{j_1}, \dots, \sigma_{j_{k'}})$  is equal to

$$\sum_{\sigma_{l_1}, \dots, \sigma_{l_r} \in \Pi(d)} g_1(\sigma)g_2(\sigma') + \sum_{\sigma_{l_1}, \dots, \sigma_{l_r} \in \Pi(d)} [e_1(\sigma)g_2(\sigma') + e_2(\sigma')g_1(\sigma) + e_1(\sigma)e_2(\sigma')]$$

The first term of this sum is simply  $Contr(g_1, g_2)(\sigma_{i_1}, \dots, \sigma_{i_k}, \sigma_{j_1}, \dots, \sigma_{j_{k'}})$ . The second term can be simplified by noting that  $|e_1(\sigma)| \leq \varepsilon$ ,  $|e_2(\sigma')| \leq \varepsilon$ ,  $|g_1(\sigma)| \leq 1$ ,  $|g_2(\sigma')| \leq 1$  and that  $|e_1(\sigma)e_2(\sigma')| \leq \varepsilon$ . Additionally, since  $\mathcal{I}_1 \cap \mathcal{I}_2 = \{l_1, \dots, l_r\}$ , there are at most  $d^{2|\mathcal{I}_1 \cap \mathcal{I}_2|}$  sequences of the form  $\sigma_{l_1}, \dots, \sigma_{l_r}$  where  $\sigma_{l_i} \in \Pi(d)$  for each  $i \in \{1, \dots, r\}$ . Therefore,  $Contr(g'_1, g'_2)(\sigma_{i_1}, \dots, \sigma_{i_k}, \sigma_{j_1}, \dots, \sigma_{j_{k'}})$  is equal to

$$Contr(g_1, g_2)(\sigma_{i_1}, \dots, \sigma_{i_k}, \sigma_{j_1}, \dots, \sigma_{j_{k'}}) \pm \varepsilon \cdot (3d^{2|\mathcal{I}_1 \cap \mathcal{I}_2|}).$$

□

We observe that for each three tensors  $g_1$ ,  $g_2$  and  $g_3$ , if  $\|g_1 - g_2\| \leq \varepsilon$  and  $\|g_2 - g_3\| \leq \varepsilon'$  then  $\|g_1 - g_3\| \leq \varepsilon + \varepsilon'$ . Using this observation, the following lemma is a consequence of Lemma 3.

**Lemma 4** *Let  $g_1$  and  $g'_1$  be tensors with index set  $\mathcal{I}_1$ , and  $g_2$  and  $g'_2$  be tensors with index set  $\mathcal{I}_2$  where  $\mathcal{I}_1 \cap \mathcal{I}_2 \neq \emptyset$ ,  $|\mathcal{I}_1| \leq r$  and  $|\mathcal{I}_2| \leq r$ . Let  $\|g_1 - g_2\| \leq \varepsilon \cdot (3d^{2 \cdot r} + 1)^h$  and  $\|g'_1 - g'_2\| \leq \varepsilon \cdot (3d^{2 \cdot r} + 1)^h$ . Then*

$$\|Trunc_\varepsilon(Contr(g_1, g_2)) - Contr(g'_1, g'_2)\| \leq \varepsilon \cdot (3d^{2 \cdot r} + 1)^{h+1}.$$

*Proof* By our definition of truncation,

$$\| \text{Trunc}_\varepsilon(\text{Contr}(g_1, g_2)) - \text{Contr}(g_1, g_2) \| \leq \varepsilon.$$

Since  $|\mathcal{I}_1| \leq r$  and  $|\mathcal{I}_2| \leq r$ , we have  $|\mathcal{I}_1 \cap \mathcal{I}_2| \leq r$ . Therefore, by Lemma 3,

$$\| \text{Contr}(g_1, g_2) - \text{Contr}(g'_1, g'_2) \| \leq \left[ \varepsilon \cdot (3d^{2 \cdot r} + 1)^h \right] \cdot 3d^{2 \cdot r}.$$

This implies that

$$\begin{aligned} \| \text{Trunc}_\varepsilon(\text{Contr}(g_1, g_2)) - \text{Contr}(g'_1, g'_2) \| &\leq \varepsilon \cdot (3d^{2 \cdot r} + 1)^h \cdot 3d^{2 \cdot r} + \varepsilon \\ &\leq \varepsilon \cdot (3d^{2 \cdot r} + 1)^{h+1}. \end{aligned}$$

□

Next, we define the notion of *partial simulation*. Recall that if  $T$  is a tree, then we denote by  $T[u]$  the subtree of  $T$  rooted at  $u$ .

**Definition 12 (Partial Simulation)** Let  $(\mathcal{N}, A)$  be a feasibility tensor network,  $(T, \iota)$  be a contraction tree for  $\mathcal{N}$  and  $u$  be a node of  $T$ . A partial simulation of  $(\mathcal{N}, A)$  rooted at  $u$  is a function  $\hat{\lambda}_u : \text{nodes}(T[u]) \rightarrow \mathbb{T}(d)$  satisfying the following conditions.

1. For each leaf  $u'$  of  $T[u]$ ,  $\hat{\lambda}_u(u') \in A(u')$ .
2. For each internal node  $u'$  of  $T[u]$ ,  $\hat{\lambda}_u(u') = \text{Contr}(\hat{\lambda}_u(u'.l), \hat{\lambda}_u(u'.r))$

Intuitively, a partial simulation  $\hat{\lambda}_u : \text{nodes}(T[u]) \rightarrow \mathbb{T}(d)$  of  $(\mathcal{N}, A)$  can be obtained by the following process. First, we consider some initialization  $\lambda$  of  $(\mathcal{N}, A)$ . Subsequently, we construct the simulation  $\hat{\lambda} : \text{nodes}(T) \rightarrow \mathbb{T}(d)$  of the tensor network  $(\mathcal{N}, \lambda)$  on the contraction tree  $(T, \iota)$ . Finally, we restrict  $\hat{\lambda}$  to the nodes of  $T[u]$ . In other words, we set  $\hat{\lambda}_u = \hat{\lambda}|_{\text{nodes}(T[u])}$ . In particular, we note that if  $u$  is the root of  $T$ , then  $\hat{\lambda}_u = \hat{\lambda}$ . The next lemma establishes an upper bound for the error propagation during the process of constructing an  $\varepsilon$ -simulation for a feasibility tensor network.

**Lemma 5** Let  $(\mathcal{N}, A)$  be a feasibility tensor network,  $(T, \iota)$  be a contraction tree for  $\mathcal{N}$  of rank  $r$  and  $\hat{A} : N \rightarrow 2^{\mathbb{T}(d, \varepsilon)}$  be an  $\varepsilon$ -simulation of  $(\mathcal{N}, A)$  on  $(T, \iota)$ . Finally, let  $u$  be a node of  $T$ .

1. For each partial simulation  $\hat{\lambda}_u : \text{nodes}(T[u]) \rightarrow \mathbb{T}(d)$  of  $(\mathcal{N}, A)$  rooted at  $u$ , there is a tensor  $g \in \hat{A}(u)$  such that  $\|\hat{\lambda}_u(u) - g\| \leq \varepsilon \cdot (3d^{2 \cdot r} + 1)^{\text{height}(u)}$ .
2. For each tensor  $g \in \hat{A}(u)$ , there is a partial simulation  $\hat{\lambda}_u : \text{nodes}(T[u]) \rightarrow \mathbb{T}(d)$  of  $(\mathcal{N}, A)$  rooted at  $u$  such that  $\|\hat{\lambda}_u(u) - g\| \leq \varepsilon \cdot (3d^{2 \cdot r} + 1)^{\text{height}(u)}$ .

*Proof* The proofs of both Lemma 5.1 and Lemma 5.2 follow by induction on the height of  $u$ . First, we note that since the contraction tree  $(T, \iota)$  has rank  $r$ , we have that  $|\mathcal{I}(u.l) \cap \mathcal{I}(u.r)| \leq r$  for each internal node  $u$  of  $T$ . Thus, all tensors associated with nodes of  $T$  have rank at most  $r$ . Now we proceed with the proof by induction. In the base case,  $u$  is a leaf and therefore,  $\text{nodes}(T[u]) = \{u\}$ . In this case, for each partial simulation  $\hat{\lambda}_u : \{u\} \rightarrow \mathbb{T}(d)$ , the tensor  $\hat{\lambda}_u(u)$  belongs to  $A(u) = \hat{A}(u)$  by Definition 12. Conversely, for each tensor  $g \in \hat{A}(u)$ , the function  $\hat{\lambda}_u : \{u\} \rightarrow \mathbb{T}(d)$  obtained by setting  $\hat{\lambda}_u(u) = g$  is a valid partial simulation. Now, suppose that the lemma is valid for every node of height at most  $h$  and let  $u$  be a node of height  $h + 1$ .

1. Let  $\hat{\lambda}_u : \text{nodes}(T[u]) \rightarrow \mathbb{T}(d)$  be a partial simulation. Let  $\hat{\lambda}_{u.l} = \hat{\lambda}_u|_{T[u.l]}$  and  $\hat{\lambda}_{u.r} = \hat{\lambda}_u|_{T[u.r]}$  be the restrictions of  $\hat{\lambda}_u$  to the nodes of the subtrees  $T[u.l]$  and  $T[u.r]$  respectively. Note that  $\hat{\lambda}_{u.l}(u.l) = \hat{\lambda}_u(u.l)$  and  $\hat{\lambda}_{u.r}(u.r) = \hat{\lambda}_u(u.r)$ , and therefore, by Definition 12,

$$\hat{\lambda}_u(u) = \text{Contr}(\hat{\lambda}_{u.l}(u.l), \hat{\lambda}_{u.r}(u.r)).$$

By the induction hypothesis, there exist tensors  $g.l \in \hat{A}(u.l)$  and  $g.r \in \hat{A}(u.r)$  such that

$$\|\hat{\lambda}_{u.l}(u.l) - g.l\| \leq \varepsilon \cdot (3d^{2r} + 1)^h$$

and

$$\|\hat{\lambda}_{u.r}(u.r) - g.r\| \leq \varepsilon \cdot (3d^{2r} + 1)^h.$$

By Definition 11, the tensor  $g = \text{Trunc}_\varepsilon(\text{Contr}(g.l, g.r))$  belongs to  $\hat{A}(u)$ . Finally, by Lemma 4,

$$\|\hat{\lambda}_u(u) - g\| \leq \varepsilon \cdot (3d^{2r} + 1)^{h+1}.$$

2. Let  $g$  be a tensor in  $\hat{A}(u)$ . By Definition 11, there exist tensors  $g.l \in \hat{A}(u.l)$  and  $g.r \in \hat{A}(u.r)$  such that  $g = \text{Trunc}_\varepsilon(\text{Contr}(g.l, g.r))$ . By the induction hypothesis, there exist partial simulations  $\hat{\lambda}_{u.l} : \text{nodes}(T[u.l]) \rightarrow \mathbb{T}(d)$  and  $\hat{\lambda}_{u.r} : \text{nodes}(T[u.r]) \rightarrow \mathbb{T}(d)$  such that

$$\|\hat{\lambda}_{u.l}(u.l) - g.l\| \leq \varepsilon \cdot (3d^{2r} + 1)^h$$

and

$$\|\hat{\lambda}_{u.r}(u.r) - g.r\| \leq \varepsilon \cdot (3d^{2r} + 1)^h.$$

Now let  $\hat{\lambda}_u : \text{nodes}(T[u]) \rightarrow \mathbb{T}(d)$  be the partial simulation that extends  $\hat{\lambda}_{u.l}$  and  $\hat{\lambda}_{u.r}$  by one node. More precisely, restricting  $\hat{\lambda}_u$  to the nodes of  $T[u.l]$  yields  $\hat{\lambda}_{u.l}$ , restricting  $\hat{\lambda}_u$  to the nodes of  $T[u.r]$  yields  $\hat{\lambda}_{u.r}$ , and the tensor associated by  $\hat{\lambda}_u$  with the node  $u$  is the contraction of the tensors associated by  $\hat{\lambda}_{u.l}$  and  $\hat{\lambda}_{u.r}$  with the nodes  $u.l$  and  $u.r$  respectively. Formally,  $\hat{\lambda}_u$  is defined by setting

$$\hat{\lambda}_u|_{T[u.r]} = \hat{\lambda}_{u.r}, \quad \hat{\lambda}_u|_{T[u.l]} = \hat{\lambda}_{u.l}, \quad \text{and}$$

$$\hat{\lambda}_u(u) = \text{Contr}(\hat{\lambda}_{u.l}(u.l), \hat{\lambda}_{u.r}(u.r)).$$

Therefore, by Lemma 4,

$$\|\hat{\lambda}_u(u) - g\| \leq \varepsilon \cdot (3d^{2r} + 1)^{h+1}.$$

□

The next lemma, which is a consequence of Lemma 5, states that if  $u_0$  is the root of  $T$  then the maximum absolute value of a complex number in  $\hat{A}(u_0)$  is at most  $(3d^{2r} + 1)^h$  apart from the value  $\text{VAL}(\mathcal{N}, A)$  of the feasibility tensor network  $(\mathcal{N}, A)$ .

**Lemma 6** *Let  $(\mathcal{N}, A)$  be a feasibility tensor network and  $(T, \iota)$  be a contraction tree for  $\mathcal{N}$  of rank  $r$  and height  $h$ . Let  $\hat{A} : N \rightarrow 2^{\mathbb{T}(d, \varepsilon)}$  be an  $\varepsilon$ -simulation of  $(\mathcal{N}, A)$  on  $(T, \iota)$ . If  $u_0$  is the root of  $T$  and  $\alpha$  is the largest absolute value of a complex number in  $\hat{A}(u_0)$ , then  $|\alpha - \text{VAL}(\mathcal{N}, A)| \leq (3d^{2r} + 1)^h$ .*

*Proof* First, we note that since  $u_0$  is the root of  $T$ , all elements of  $\hat{A}(u_0)$  are complex numbers, i.e., rank-0 tensors.

*Claim* For each initialization  $\lambda$  of  $(\mathcal{N}, A)$  there is a complex number in  $\hat{A}(u_0)$  such that  $|\text{val}(\mathcal{N}, \lambda) - |g|| \leq (3d^{2^r} + 1)^h$ . Conversely, for each  $g \in \hat{A}(u_0)$  there is an initialization  $\lambda$  of  $(\mathcal{N}, A)$  such that  $|\text{val}(\mathcal{N}, \lambda) - |g|| \leq (3d^{2^r} + 1)^h$ .

Recall that  $\text{VAL}(\mathcal{N}, A)$  is defined as  $\max_{\lambda} \text{val}(\mathcal{N}, \lambda)$  where  $\lambda$  ranges over all initializations of  $(\mathcal{N}, A)$ . Therefore, the claim stated above implies that if  $\alpha$  is the maximum absolute value of a complex number in  $\hat{A}(u_0)$ , then

$$|\alpha - \text{VAL}(\mathcal{N}, A)| \leq (3d^{2^r} + 1)^h.$$

Now we proceed to prove our claim. First, let  $\lambda$  be an initialization of  $(\mathcal{N}, A)$ . Then there exists a partial simulation  $\hat{\lambda}_{u_0}$  of  $(\mathcal{N}, A)$  rooted at  $u_0$  such that  $\hat{\lambda}_{u_0} = \hat{\lambda}$ , where  $\hat{\lambda}$  is the simulation of the tensor network  $(\mathcal{N}, \lambda)$  constructed as in Definition 7. Note that  $\text{val}(\mathcal{N}, \lambda) = |\hat{\lambda}_{u_0}(u_0)|$ , since  $u_0$  is the root of  $T$ . By Lemma 5.i, there exists a complex number  $g \in \hat{A}(u_0)$  such that  $|\hat{\lambda}_{u_0}(u_0) - g| \leq (3d^{2^r} + 1)^h$ . Using the fact that  $||x| - |y|| \leq |x - y|$  for every pair of complex numbers  $x$  and  $y$ , we have that  $|\text{val}(\mathcal{N}, \lambda) - |g|| \leq (3d^{2^r} + 1)^h$ .

Conversely, let  $g$  be a complex number in  $\hat{A}(u_0)$ . By Lemma 5.ii, there exists a partial simulation  $\lambda_{u_0}$  of  $(\mathcal{N}, A)$  rooted at  $u_0$  such that  $|\lambda_{u_0}(u_0) - g| \leq (3d^{2^r} + 1)^h$ . Since  $u_0$  is the root of  $T$ , there exists an initialization  $\lambda$  of  $(\mathcal{N}, A)$  such that  $\hat{\lambda}_{u_0} = \hat{\lambda}$ , where  $\hat{\lambda}$  is the simulation of the tensor network  $(\mathcal{N}, \lambda)$  on  $(T, \iota)$  constructed according to Definition 7. Note again that  $\text{val}(\mathcal{N}, \lambda) = |\hat{\lambda}_{u_0}(u_0)|$ . Therefore using the fact that  $||x| - |y|| \leq |x - y|$  for every pair of complex numbers  $x$  and  $y$ , we have that  $|\text{val}(\mathcal{N}, \lambda) - |g|| \leq (3d^{2^r} + 1)^h$ . This proves the claim.  $\square$

Finally, we are in a position to prove Theorem 5.

### Proof of Theorem 5

1) Let  $(\mathcal{N}, A)$  be a feasibility tensor network and  $(T, \iota)$  be a contraction tree for  $\mathcal{N}$  of rank  $r$  and height  $h$ . Let  $\hat{A} : N \rightarrow 2^{\mathbb{T}(d)}$  be the  $\varepsilon$ -simulation  $(\mathcal{N}, A)$  on  $(T, \iota)$  constructed according to Definition 11. Since  $T$  is a binary tree with  $|\mathcal{N}|$  leaves, the total number of nodes in  $T$  is  $2|\mathcal{N}| - 1$ . Additionally, since  $(T, \iota)$  has rank  $r$ , for each node  $u$  of  $T$ , the set  $\hat{A}(u)$  has at most  $|\mathbb{T}(d, \varepsilon, r)| = \varepsilon^{-\exp(O(r \cdot \log d))}$  tensors. Therefore,  $\hat{A}$  can be constructed in time  $|\mathcal{N}| \cdot \varepsilon^{-\exp(O(r \cdot \log d))}$ . Now let  $u_0$  be the root of  $T$ , and let  $\alpha$  be the largest absolute value of a complex number in  $\hat{A}(u_0)$ . Then by Lemma 6,  $|\alpha - \text{VAL}(\mathcal{N}, A)| \leq (3d^{2^r} + 1)^h$ . Therefore, after having constructed  $\hat{A}$  we just need to output  $\alpha$ .

2) Let  $u_0$  be the root of  $T$  and let  $g$  be the complex number in  $\hat{A}(u_0)$  with largest absolute value  $\alpha$ . We construct a simulation  $\hat{\lambda} : \text{nodes}(T) \rightarrow \mathbb{T}(d)$  of  $(\mathcal{N}, A)$  on  $(T, \iota)$  as follows. First, we set  $\hat{\lambda}(u_0) = g$ . Now for each internal node  $u$  for which  $\hat{\lambda}(u)$  has already been determined, let  $u_l$  and  $u_r$  be respectively the left and right children of  $u$ . Then we set  $\hat{\lambda}(u.l) = g_l$  and  $\hat{\lambda}(u.r) = g_r$  where  $g_l$  and  $g_r$  are respectively tensors in  $\hat{A}(u.l)$  and  $\hat{A}(u.r)$  for which  $\hat{\lambda}(u) = \text{Trunc}_{\varepsilon}(g_l, g_r)$ . We proceed in this way until we have determined  $\hat{\lambda}$  on all leaves of  $T$ . The searched initialization  $\lambda$  is then obtained by considering the tensors associated by  $\hat{\lambda}$  with the leaves of  $T$ . In other words, for each leaf  $u$  labeled with the index set  $\iota(u)$ , we set  $\lambda(\iota(u)) = \hat{\lambda}(u)$ . Since  $\hat{\lambda}(u_0) = g$ , we have that  $\text{val}(\mathcal{N}, \lambda) = \alpha$ . Since  $T$  has



$2|\mathcal{N}| - 1$  nodes, once we are given  $\hat{\Lambda}$ , the construction of the initialization  $\lambda$  takes time  $O(|\mathcal{N}|)$ . Since  $\hat{\Lambda}$  can be constructed in time  $|\mathcal{N}| \cdot \varepsilon^{-\exp(O(r \cdot \log d))}$ , the overall time complexity to construct  $\lambda$  is  $|\mathcal{N}| \cdot \varepsilon^{-\exp(O(r \cdot \log d))}$ .  $\square$

## 8 Classical Witnesses vs Quantum Verifiers of Logarithmic Width

In this Section we will prove Theorem 2, which states that for any constant  $\delta$  with  $0 < \delta < 1$ , and any quantum circuit  $C$  with  $n$  uninitialized inputs,  $\text{poly}(n)$  gates, and online-width  $O(\log n)$ , it is NP complete to determine whether  $\text{Pr}^{cl}(C) = 1$  or whether  $\text{Pr}^{cl}(C) \leq \delta$ . We note that as an implication of Theorem 2, we have that  $\text{QCMA}[\text{ow}, O(\log n)] = \text{NP}$ . Indeed, from the proof of Theorem 2, it will also follow that  $\text{QCMA}[\text{tw}, O(\log n)] = \text{NP}$ .

**Membership in NP:** As mentioned in the introduction, Markov and Shi's simulation algorithm [20] computes (with polynomially many bits of precision) the acceptance probability of a quantum circuit of treewidth  $t$  in deterministic time  $2^{O(t)} \cdot |C|^{O(1)}$ . Let  $C$  be a quantum circuit with  $n$  uninitialized inputs,  $\text{poly}(n)$  gates, and treewidth  $O(\log n)$ . Given a classical assignment  $y \in \{0, \dots, d-1\}^n$  we can use Markov and Shi's algorithm to compute  $\text{Pr}(C, |y\rangle)$  in time  $n^{O(1)}$ . We accept  $y$  as a suitable witness if  $\text{Pr}(C, |y\rangle) = 1$  and we reject  $y$  if  $\text{Pr}(C, |y\rangle) \leq \delta$ . This shows that the problem of computing a  $\delta$ -optimal classical witness for a quantum circuit of logarithmic treewidth is in NP. Since, by Equation 1, the treewidth of any circuit is upper bounded by its online width, we have that the problem of computing a  $\delta$ -optimal classical witness for a quantum circuit of logarithmic online-width is also in NP. In summary, we have just shown that

$$\text{QCMA}[\text{ow}, O(\log n)] \subseteq \text{QCMA}[\text{tw}, O(\log n)] \subseteq \text{NP}.$$

**NP-hardness:** To show that the problem of finding a  $\delta$ -optimal classical assignment for a quantum circuit of logarithmic online-width is NP-hard, we will first consider a probabilistic verifier of logarithmic online-width for the 3-SAT problem which has inverse polynomial soundness. Subsequently we will show that the soundness can be amplified to a constant by increasing the online-width of the original verifier by a logarithmic additive factor. Clearly, such hardness result for probabilistic circuits of logarithmic online width implies the same hardness result for quantum circuits of logarithmic online width. Also, since the treewidth of a circuit is always upper-bounded by its online-width, the following sequence of inclusions is implied by the NP-hardness proof provided in this section.

$$\text{NP} \subseteq \text{QCMA}[\text{ow}, O(\log n)] \subseteq \text{QCMA}[\text{tw}, O(\log n)].$$

**Proposition 5 (Folklore)** *Let  $F(x_1, x_2, \dots, x_n)$  be a 3-SAT formula with  $n$  variables  $x_1 \dots x_n$  and  $m = \text{poly}(n)$  clauses  $W_1, W_2, \dots, W_m$ . There is a probabilistic circuit  $C_F$  of online-width  $O(\log n)$  such that the following conditions are satisfied.*

1. (Completeness) *If  $F$  is satisfiable then  $\text{Pr}^{cl}(C_F) = 1$ .*
2. (Soundness) *If  $F$  is not satisfiable then  $\text{Pr}^{cl}(C_F) \leq 1 - 1/\text{poly}(n)$ .*

*Proof* The circuit  $C_F$  simulates the execution of an online verifier that does the following: First it chooses a number  $r$  from the set  $\{1, \dots, m\}$  at random. Then, when reading the witness provided by Merlin, it ignores all bits which assign values to variables that do not belong to  $W_r$ , and keeps only those bits which correspond to variables used by  $W_r$ . Subsequently it verifies whether  $W_r$  is evaluated to true. If the formula  $F$  has a satisfying assignment, then the prover can always convince the verifier with probability 1, since in this case the restriction of a satisfying assignment to the variables occurring in  $W_r$  will cause  $W_r$  to evaluate to 1. Otherwise, if  $F$  has no satisfying assignment, then no matter what assignment is provided by the prover, there will be at least one clause of  $F$  which will not be satisfied. Thus the verifier will accept with probability at most  $1 - 1/m$ .  $\square$

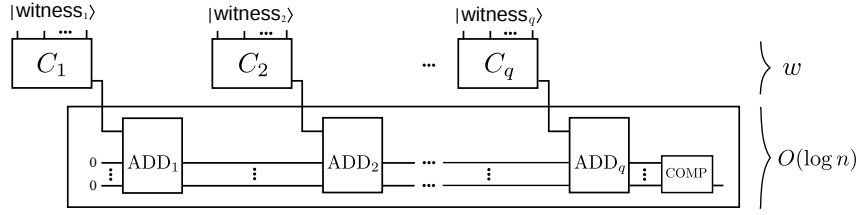
It remains to show that the same hardness result holds if the soundness in Proposition 5 is constant, instead of inverse polynomial. This can be done by a standard probability amplification argument: First we create polynomially many copies of the original verifier, each with access to independent random bits, and then we consider the majority vote of the answer of all verifiers. We claim that if the original verifier  $C$  has online-width  $w$ , then the overall amplified circuit has online-width  $w + O(\log n)$ .

Let  $q = n^{O(1)}$ , and let ADD be an adder with  $c + 1$  input bits and  $c$  output bits where  $c = \lceil \log q \rceil$ . One of the inputs of ADD is a control bit. The other  $c$  inputs of ADD form a counter register. The circuit ADD increases the value of its counter register by one if the control bit is set to 1, and does nothing with the input otherwise. Let COMP be a circuit with  $c$  input bits and one output bit. The circuit COMP accepts if the value represented by its inputs is greater than  $\lceil \frac{q}{2} \rceil$ , and rejects otherwise. Finally let MAJ( $x_1, \dots, x_q$ ) be a circuit that accepts an input  $x_1 x_2 \dots x_q$  if  $\sum_{i=1}^q x_i \geq \lceil \frac{q}{2} \rceil$  and rejects otherwise. Then MAJ( $x_1, \dots, x_q$ ) can be implemented as a sequence of sub-circuits ADD<sub>1</sub>, ADD<sub>2</sub>, ..., ADD<sub>q</sub>, COMP, where for each  $i \in \{1, \dots, q\}$ , ADD <sub>$i$</sub>  is an adder with control bit  $x_i$ , for each  $i \in \{1, \dots, q-1\}$  the outputs of ADD <sub>$i$</sub>  are identified with the inputs at the counter register of ADD <sub>$i+1$</sub> , the outputs of ADD <sub>$q$</sub>  are identified with the inputs of COMP, and the inputs at the counter register of ADD<sub>1</sub> are set to 0 (Fig. 5).

Now let  $C_F$  be the probabilistic circuit of online width  $w = O(\log n)$  considered in Proposition 5. Let  $C_1, \dots, C_q$  be  $q$  independent copies of the circuit  $C_F$ . Finally, let  $C' = \text{MAJ}(C_1, C_2, \dots, C_q)$  be the circuit obtained from MAJ( $x_1, \dots, x_q$ ) by identifying, for each  $i \in \{1, \dots, q\}$ , the output bit of  $C_i$  with the control bit of the sub-circuit ADD <sub>$i$</sub> . Then  $C'$  accepts if and only if at least  $\lceil \frac{q}{2} \rceil$  of the circuits  $C_i$  accept. For each constant  $\delta$  with  $0 < \delta < 1$ , we can choose a suitable  $q = n^{O(1)}$ , such that  $C'$  accepts with probability 1 if the 3-SAT formula  $F$  is satisfiable, and accepts with probability at most  $\delta$  if  $F$  is unsatisfiable. To show that  $C'$  has online-width at most  $w + O(\log n)$ , we note that there is a topological ordering of the gates of  $C'$  which executes all gates of  $C_i$  before all gates of ADD <sub>$i$</sub> , all gates of ADD <sub>$i$</sub>  before all gates of  $C_{i+1}$ , and all gates of ADD <sub>$q$</sub>  before all gates of COMP (Fig. 5).  $\square$

## 9 Quantum Witnesses vs Quantum Verifiers of Logarithmic Width

In this section we will prove Theorem 3, which states that for any constant  $\delta$  with  $0 < \delta < 1/2$ , and any quantum circuit  $C$  with  $n$  uninitialized inputs,



**Fig. 5** Amplification of Probability with a logarithmic increase in online-width. The sub-circuit surrounded by the box implements the majority vote in online-width  $O(\log n)$ .

$\text{poly}(n)$  gates and online-width  $O(\log n)$ , it is QMA-Complete to determine whether  $\text{Pr}^{qu}(C) \geq 1 - \delta$  or whether  $\text{Pr}^{qu}(C) \leq \delta$ . We note that membership in QMA is trivial, since QMA is defined in terms of the quantum satisfiability of quantum circuits of polynomial online-width. On the other hand, the proof of QMA-hardness will be similar to the proof of NP-hardness for the classical satisfiability of quantum circuits of logarithmic online-width given in Section 8. The only difference is that instead of using a reduction from 3-SAT, we will use a reduction from the QMA-Complete problem  $k$ -local Hamiltonian [19]. We note that this completeness result implies that  $\text{QMA}[\text{ow}, O(\log n)] = \text{QMA}$ . Since, by Equation 1, the treewidth of a circuit is always upper bounded by its online width, we also have that  $\text{QMA}[\text{tw}, O(\log n)] = \text{QMA}$ .

Let  $\mathcal{H}_d = \mathbb{C}^d$ . An operator  $H : \mathcal{H}_d^{\otimes n} \rightarrow \mathcal{H}_d^{\otimes n}$  is called a  $k$ -local Hamiltonian if it is expressible in the form  $H = \sum_j H_j$  where each  $H_j$  is an Hermitian operator acting on at most  $k$  qubits. Additionally we assume a normalizing condition requiring both  $H_j$  and  $I - H_j$  to be positive semidefinite.

**Definition 13 ( $k$ -Local Hamiltonian Problem)** Let  $k = O(1)$ ,  $H = \sum_{i=1}^m H_i$  be a  $k$ -local Hamiltonian acting on  $n$  qubits, and  $a, b$  be real numbers such that  $0 \leq a < b$  and  $b - a = \Omega(n^{-\alpha})$  for some constant  $\alpha > 0$ . The  $k$ -local Hamiltonian problem consists in determining whether  $H$  has an eigenvalue not exceeding  $a$ , or whether all eigenvalues of  $H$  are at least  $b$ .

The  $k$ -local Hamiltonian problem was shown to be QMA complete in [19]. In particular, the proof that this problem is in QMA follows from a reduction to the problem of approximating the maximum acceptance probability of a quantum circuit with uninitialized inputs. Next, we show that this reduction carry over even if we require the obtained quantum circuits to have logarithmic online-width. We note that the exposition given below is similar to the one encountered in [19], except for some adaptations that take the online-width of the involved circuits into consideration.

**Lemma 7 (Adaptation from [19])** Let  $k = O(1)$ ,  $H = \sum_{i=1}^m H_i$  be a  $k$ -local Hamiltonian acting on  $n$  qubits, and  $a, b$  be real numbers such that  $0 \leq a < b$  and  $b - a = \Omega(n^{-\alpha})$  for some constant  $\alpha > 0$ . Then there is a quantum circuit  $C_H$  with  $n$  uninitialized inputs and online-width  $O(\log n)$  satisfying the following conditions.

1. (Completeness) If some eigenvalue of  $H$  is smaller than or equal to  $a$ , then  $\text{Pr}^{qu}(C_H) \geq 1 - m^{-1}a$ .
2. (Soundness) If all eigenvalues of  $H$  are at least  $b$ , then  $\text{Pr}^{qu}(C_H) \leq 1 - m^{-1}b$ .

*Proof* Let  $H = \sum_{j=1}^m H_j$  be a  $k$ -local Hamiltonian. For each local term  $H_j$  we construct a circuit implementing the POVM  $\{H_j, I - H_j\}$ . Since  $H_j$  can be rewritten as  $H_j = \sum_s \lambda_s |\psi_s\rangle\langle\psi_s|$ , where  $|\psi_s\rangle$  are the eigenvectors of  $H_j$ , and since  $H_j$  acts on a constant number of qubits, the mentioned POVM can be implemented by a constant size circuit  $W_j$  that acts on the qubits affected by  $H_j$  and an auxiliary output qubit. The action of  $W_j$  on the orthogonal system of eigenvectors of  $H_j$  is given by

$$W_j : |\psi_s, 0\rangle \rightarrow |\psi_s\rangle \otimes (\sqrt{\lambda_s}|0\rangle + \sqrt{1 - \lambda_s}|1\rangle).$$

The probability of measuring 1 at the output bit of  $W_j$  is given by

$$Pr_1(W_j) = \langle\eta, 0| W_j^\dagger (I \otimes |1\rangle\langle 1|) W_j |\eta, 0\rangle = 1 - \langle\eta|H_j|\eta\rangle.$$

Now consider a circuit  $C_H$  which implements the following verification process. First, the verifier selects a number  $r \in \{1, \dots, m\}$  uniformly at random. Subsequently, when reading the witness  $|\eta\rangle$  provided by Merlin, the verifier ignores all qubits but those which are affected by  $W_r$ . Finally when all relevant qubits have been read, the verifier applies the sub-circuit  $W_r$  to these relevant qubits. The overall acceptance probability of the circuit  $C_H$  is given by

$$Pr(C_H, |\eta\rangle) = \sum_j \frac{1}{r} Pr(W_j, |\eta\rangle) = 1 - m^{-1} \langle\eta|H|\eta\rangle.$$

In particular, if  $|\eta\rangle$  is an eigenvector of  $H$  with eigenvalue smaller than  $a$ , then the acceptance probability of  $C_H$  is greater than  $1 - m^{-1}a$  while if every eigenvector of  $H$  has eigenvalue at least  $b$  then the acceptance probability of  $C_H$  is at most  $1 - m^{-1}b$ . The circuit  $C_H$  can clearly be implemented in online-width  $O(\log n)$  since we just need  $O(\log n)$  bits to implement the random choice of  $r$ .  $\square$

To prove Theorem 3, it remains to show that both the soundness and the completeness in Lemma 7 can be amplified with only a logarithmic increase in online-width. Let  $q = n^{O(1)}$ ,  $p(n) = (a(n) + b(n))/2$ , and let  $\text{MAJ}'(x_1, \dots, x_q)$  be a circuit that implements the following variant of the majority function.

$$\text{MAJ}'(x_1, \dots, x_q) = \begin{cases} 1 & \text{if } \sum_{j=1}^q x_j \geq p(n) \cdot q \\ 0 & \text{if } \sum_{j=1}^q x_j < p(n) \cdot q \end{cases} \quad (15)$$

Let  $C_1, \dots, C_q$  be independent copies of the circuit  $C_H$ . Let  $C' = \text{MAJ}'(C_1, \dots, C_q)$  be the circuit obtained from  $\text{MAJ}'(x_1, \dots, x_q)$  by identifying the output of  $C_i$  with the  $i$ -th input of  $\text{MAJ}'(x_1, \dots, x_q)$ . It can be shown (See [19] Lemma 14.1) that if there exists a witness  $|\eta\rangle \in \mathcal{H}_d^{\otimes n}$  such that  $Pr(C_H, |\eta\rangle) \geq a(n)$ , then there is a witness  $|\eta'\rangle \in \mathcal{H}_d^{\otimes qn}$  such that

$$Pr(C', |\eta'\rangle) \geq 1 - \exp(-\Omega(\text{poly}(n))).$$

On the other hand, if for every state  $|\eta\rangle \in \mathcal{H}_d^{\otimes n}$ ,  $Pr(C_H, |\eta\rangle) \leq b(n)$ , then for every state  $|\eta'\rangle \in \mathcal{H}_d^{\otimes qn}$ , the verifier accepts with probability at most  $\exp(-\text{poly}(n))$ . Similarly to the circuit computing  $\text{MAJ}(x_1, \dots, x_q)$  described in Section 8, the circuit  $\text{MAJ}'(x_1, \dots, x_q)$  can be implemented in  $O(\log_n)$  as a sequence

$$\text{ADD}_1, \text{ADD}_2, \dots, \text{ADD}_q, \text{COMP}'$$

of adder circuits followed by a comparator circuit  $\text{COMP}'$  which accepts if and only if the value at its input register is at least  $p(n) \cdot q$ . As in Section 8, this implies that the overall circuit  $C'$  has online-width at most  $w + O(\log n)$ , since we can consider an ordering of the gates of  $C'$  that executes all gates of  $C_i$  before the gates of the adder circuit  $\text{ADD}_i$ , all gates of  $\text{ADD}_i$  before the gates of  $C_{i+1}$ , and all gates of  $\text{ADD}_q$  before all gates of  $\text{COMP}'$  (Fig. 5).  $\square$

## 10 Conclusion and Open Problems

In this work we have introduced the notion of *feasibility tensor network*. We have shown that the problem of computing a classical assignment  $y \in \{0,1\}^n$  that maximizes the acceptance probability of a quantum circuit  $C$  with  $n$  uninitialized inputs and  $\text{poly}(n)$  gates can be reduced to the problem of finding an initialization of maximum value for a feasibility tensor network. Using this reduction, we have shown that if  $C$  has treewidth  $t$ , then a  $\delta$ -optimal assignment for  $C$  can be found in time  $(n/\delta)^{\exp(O(t))}$ . Therefore we have provided the first example of quantum optimization problem that can be solved in polynomial time on quantum circuits of constant treewidth.

We have also provided new characterizations of the complexity classes NP and QMA in terms of Merlin-Arthur protocols in which the verifier is a circuit of logarithmic treewidth, by showing that  $\text{QCMA}[\text{tw}, O(\log n)] = \text{NP}$  and that  $\text{QMA}[\text{tw}, O(\log n)] = \text{QMA}$ . In other words, we have shown that quantum witnesses are inherently more powerful than classical witnesses for Merlin-Arthur protocols with verifiers of logarithmic treewidth, assuming  $\text{QMA} \neq \text{NP}$ . Our main theorem implies that  $\text{QCMA}[\text{tw}, O(1)] \subseteq \text{P}$ . However we were not able to determine whether an analog inclusion can be proved when the verifier has constant width and the witness is allowed to be an arbitrary quantum state. More precisely, the following question is left open: Is  $\text{QMA}[\text{tw}, O(1)] \subseteq \text{P}$ ?

The NP-hardness of the problem of computing optimal classical assignments for quantum circuits of logarithmic treewidth imposes some constraints on the possibility of drastically improving the running time of our algorithm. However we leave the following question open: Is the problem of computing  $\delta$ -optimal classical assignments for quantum circuits in FPT with respect to treewidth? More precisely, can this problem be solved in time  $f(t) \cdot \text{poly}(n, \delta)$ ? We observe that while in the case of classical circuits one can determine the existence of a satisfying assignment in time  $2^{O(t)} \cdot n^{O(1)}$  [3, 4], the fact that  $\text{QCMA}[\text{tw}, O(\log n)] = \text{NP}$  implies that in the case of quantum circuits the function  $f(t)$  should be at least double exponential in  $t$ , assuming the exponential time hypothesis (ETH) [17].

## References

1. D. Aharonov, A. Kitaev, and N. Nisan. Quantum circuits with mixed states. In *Proc. of the 30th Symposium on Theory of Computing*, pages 20–30, 1998.
2. D. Aharonov and T. Naveh. Quantum NP - A survey. *arXiv preprint quant-ph/0210077*, 2002.
3. M. Alekhnovich and A. A. Razborov. Satisfiability, branch-width and Tseitin tautologies. In *Proc. of the 43rd Symposium on Foundations of Computer Science*, pages 593–603, 2002.
4. E. Allender, S. Chen, T. Lou, P. A. Papakonstantinou, and B. Tang. Width-parametrized SAT: Time-space tradeoffs. *Theory of Computing*, 10(12):297–339, 2014.

5. S. Arnborg, J. Lagergren, and D. Seese. Easy problems for tree-decomposable graphs. *Journal of Algorithms*, 12(2):308–340, 1991.
6. S. Arnborg and A. Proskurowski. Linear time algorithms for NP-hard problems restricted to partial  $k$ -trees. *Discrete Applied Mathematics*, 23(1):11–24, 1989.
7. L. Babai. Bounded round interactive proofs in finite groups. *SIAM Journal on Discrete Mathematics*, 5(1):88–111, 1992.
8. H. L. Bodlaender. Classes of graphs with bounded treewidth. *Bulletin of the EATCS*, 36:116–126, 1988.
9. H. L. Bodlaender. NC-algorithms for graphs with small treewidth. In *Proc. of the 14th International Workshop on Graph-Theoretic Concepts in Computer Science*, volume 344 of *LNCS*, pages 1–10, Springer, 1989.
10. H. L. Bodlaender, F. V. Fomin, A. M. Koster, D. Kratsch, and D. M. Thilikos. On exact algorithms for treewidth. *ACM Transactions on Algorithms*, 9(1):12, 2012.
11. A. D. Bookatz. QMA-complete problems. *Quantum Information & Computation*, 14(5-6):361–383, 2014.
12. E. Broering and S. V. Lokam. Width-based algorithms for SAT and CIRCUIT-SAT. In *Proc. of the 6th International Conference on Theory and Applications of Satisfiability Testing*, volume 2919 of *LNCS*, pages 162–171. Springer, 2004.
13. B. Courcelle. The monadic second-order logic of graphs I. Recognizable sets of finite graphs. *Information and computation*, 85(1):12–75, 1990.
14. M. de Oliveira Oliveira. On the satisfiability of quantum circuits of small treewidth. In *Proc. of the 10th International Computer Science Symposium in Russia*, volume 9139 of *LNCS*, pages 157–172, Springer, 2015.
15. K. Georgiou and P. A. Papakonstantinou. Complexity and algorithms for well-structured  $k$ -SAT instances. In *Proc. of the 11th International Conference on Theory and Applications of Satisfiability Testing*, volume 4996 of *LNCS*, pages 105–118, Springer, 2008.
16. D. Gottesman. The Heisenberg representation of quantum computers. *arXiv preprint quant-ph/9807006*, 1998.
17. R. Impagliazzo and R. Paturi. On the complexity of  $k$ -SAT. *Journal of Computer and System Sciences* 62(2):367–375, 2001.
18. R. Jozsa and N. Linden. On the role of entanglement in quantum-computational speed-up. *Proc. of the Royal Society of London, Series A*, 459(2036):2011–2032, 2003.
19. A. Kitaev, A. Shen, and M. Vyalı. *Classical and Quantum Computation*, volume 47 of *Graduate Studies in Mathematics*. AMS, 2002.
20. I. L. Markov and Y. Shi. Simulating quantum computation by contracting tensor networks. *SIAM Journal on Computing*, 38(3):963–981, 2008.
21. M. A. Nielsen and I. L. Chuang. *Quantum computation and quantum information*. Cambridge university press, 2010.
22. N. Robertson and P. D. Seymour. Graph minors III. Planar tree-width. *Journal of Combinatorial Theory, Series B*, 36(1):49–64, 1984.
23. N. Robertson and P. D. Seymour. Graph minors XIII. The disjoint paths problem. *Journal of Combinatorial Theory, Series B*, 63(1):65–110, 1995.
24. D. M. Thilikos, M. J. Serna, and H. L. Bodlaender. Constructive linear time algorithms for small cutwidth and carving-width. In *Proc. of the 11th International Conference on Algorithms and Computation*, volume 1969 of *LNCS*, pages 192–203, Springer, 2000.
25. L. G. Valiant. Quantum circuits that can be simulated classically in polynomial time. *SIAM Journal on Computing*, 31(4):1229–1254, 2002.
26. G. Vidal. Efficient classical simulation of slightly entangled quantum computations. *Physical Review Letters*, 91:147902, Oct 2003.
27. J. Watrous. Succinct quantum proofs for properties of finite groups. In *Proc. of the 41st Symposium on Foundations of Computer Science*, pages 537–546, 2000.